Independent University **Bangladesh (IUB)**

IUB Academic Repository

Computer Science and Engineering | Undergraduate Thesis

2026-08

Multi-resolution feature integration framework with CNN for Bengali handwriting quality assessment

Mondal, Shamik

Independent University, Bangladesh (IUB)

https://ar.iub.edu.bd/handle/11348/1592

Downloaded from IUB Academic Repository



Multi-Resolution Feature Integration Framework with CNN for Bengali Handwriting Quality Assessment

August 2026

Prepared by:

Shamik Mondal
ID: 2221145

Sadman Sakib
ID: 2221929

Meraj Ahmed Mozumder
ID: 2221557

Department of Computer Science and Engineering
Independent University, Bangladesh

Supervised by
Dr. Md. Tarek Habib
Associate Professor
Department of Computer Science and Engineering
Independent University, Bangladesh

Attestation

We are fully aware that plagiarism is strictly prohibited and that it violates our university's rules and regulations. We guarantee the originality and authenticity of our work. Where copyrighted materials, models, or other resources created by others have been used in our project, they have been properly acknowledged and cited in accordance with internationally recognized academic standards and guidelines.

Author Name: Shamik Mondal

.....

Signature:

.....

Author Name: Sadman Sakib

.....

Signature:

.....

Author Name: Meraj Ahmed Mozumder

.....

Signature:

.....

Letter of Transmittal

August 2026

To
Md. Tarek Habib, PhD
Associate Professor
Department of Computer Science and Engineering
Independent University, Bangladesh

Subject: Submission of Undergraduate Thesis

Dear Sir,

We are pleased to submit our undergraduate thesis entitled **"Multi-Resolution Feature Integration Framework with CNN for Bengali Handwriting Quality Assessment"** as part of the requirements for the degree of **Bachelor of Science in Computer Science and Engineering** at Independent University, Bangladesh.

Our study proposes a multi-resolution CNN-based framework for Bengali handwriting quality assessment. By integrating fine, mid, and coarse-level feature extraction, the model effectively captures complex script characteristics. The architecture outperforms several transfer learning baselines in classification tasks. It achieves an overall accuracy of **95.5%**, demonstrating strong reliability and efficiency. These results validate the effectiveness of multi-scale feature learning for real-world handwriting evaluation.

It is our utmost pleasure to express our sincere gratitude for your continuous mentorship, supervision, and support throughout the course of this research. Your invaluable guidance, constructive feedback, and encouragement have been instrumental in the successful completion of this thesis.

Thank you very much for your time, attention, and valuable support.

Sincerely,

Shamik Mondal
Sadman Sakib
Meraj Ahmed Mozumder

Attestation ..... **i**
Letter of Transmittal ..... **ii**
Abstract **1**
1. Introduction **2**
1.1 Background 2
1.2 Bengali Handwriting and Script Characteristics 3
1.2.1 Basic Characters, Modifiers and Compound Characters 3
1.2.2 Matra, Curvature and Structural Complexity 4
1.3 Handwriting Quality Assessment 5
1.4 Challenges of Bengali Handwriting Quality Assessment 5
1.4.1 Inter-Writer and Intra-Writer Variability.Inter-Writer and Intra-Writer Variability 6
1.4.2 Similar Character Shapes and Structural Ambiguity 6
1.5 Motivation of the Study 7
1.6 Problem Statement 8
1.7 Research Objectives 9
1.7.1 General Objective 9
1.7.2 Specific Objectives 9
1.8 Research Contributions 10
1.9 Scope and Limitations 10
1.9.1 Scope of the Study 10
1.9.2 Limitations of the Study 11
2. Literature Review **13**
2.1 Overview of Bengali Handwriting Research 13
2.2 Bengali Digit and Character Recognition 15
2.2.1 Handwritten Digit Recognition 16
2.2.2 Isolated Character Recognition 17
2.3 Bengali Word and Text Recognition 18
2.3.1 Word-Level Recognition 18
2.3.2 Unconstrained Handwriting Recognition 19
2.4 Bengali Handwriting Datasets 20
2.4.1 CMATERdb 20
2.4.2 Ekush and BanglaLekha-Isolated 21
2.4.3 NumtaDB and Bengali.AI Grapheme Dataset 22
2.5 CNN and Deep Learning Approaches 23
2.5.1 Custom CNN Architectures 23
2.5.2 Transfer Learning Architectures 24
2.5.3 Sequence-Based and Hybrid Approaches to Programming 24
2.6 Handwriting Quality Assessment Research 25
2.7 Comparative Analysis of Previous Studies 27
2.8 Limitations of Existing Research 29
2.9 Research Gap 30

3. Theoretical Foundations..... 32
3.1 Digital Image Processing..... 32
3.1.1 Digital Image Representation..... 32
3.1.2 Image Normalization and Resizing..... 33
3.2 Convolutional Neural Networks..... 33
3.2.1 Convolution and Feature Maps..... 34
3.2.2 Activation and Pooling..... 34
3.2.3 Batch Normalization and Dropout..... 35
3.2.4 Global Average Pooling and Softmax..... 36
3.3 Multi-Resolution Feature Learning..... 36
3.3.1 Fine-Grained Features..... 36
3.3.2 Mid-Level Structural Features..... 37
3.3.3 Global Shape Features..... 37
3.4 Feature Fusion..... 37
3.5 Transfer Learning..... 38
3.6 Optimization and Loss Function..... 39
3.6.1 Adam Optimizer..... 39
3.6.2 Categorical Cross-Entropy..... 40
3.7 Evaluation Metrics..... 40
3.7.1 Accuracy, Precision and Recall..... 40
3.7.2 F_1-Score..... 41
3.7.3 Confusion Matrix..... 41
4. Dataset Preparation and Preprocessing..... 43
4.1 Dataset Overview..... 43
4.2 Data Collection Process..... 43
4.2.1 Collection Sources..... 45
4.2.2 Collection Specifications..... 45
4.3 Handwriting Quality Classes..... 45
4.3.1 Very Bad and Bad..... 45
4.3.2 Average..... 46
4.3.3 Good and Excellent..... 46
4.4 Annotation and Quality Assessment Criteria..... 46
4.4.1 Stroke Quality..... 47
4.4.2 - Shape and Structural Correctness..... 47
4.4.3 Overall legibility and consistency..... 48
4.5 Dataset Distribution and Split..... 48
4.5.1 Training Set..... 48
4.5.2 Validation and Test Sets..... 48
4.6 Image Preprocessing..... 49
4.6.1 Image Resizing..... 49
4.6.2 Pixel Normalization..... 50

4.6.3 Grayscale-to-RGB Conversion..... 50
4.7 Data Augmentation..... 50
4.7.1 Rotation..... 52
4.7.2 Zoom and Flipping..... 52
5. Proposed Methodology..... 53
5.1 Overall Research Framework..... 53
5.2 Dataset Description and Preprocessing..... 54
5.2.1 Dataset Splitting Strategy..... 55
5.2.2 Preprocessing Pipeline..... 55
5.3 Proposed Multi-Resolution CNN Architecture..... 56
5.4 Fine-Grained Feature Stream..... 57
5.4.1 Stroke and Edge Feature Extraction..... 57
5.5 Mid-Level Feature Stream..... 58
5.5.1 Curvature and Junction Feature Extraction..... 58
5.6 Coarse Feature Stream..... 58
5.6.1 Global Shape and Spatial Layout..... 58
5.7 Multi-Resolution Feature Fusion..... 59
5.7.1 Feature Concatenation..... 59
5.7.2 Global Average Pooling..... 59
5.8 Classification Layer..... 60
5.9 Baseline Models..... 61
5.9.1 EfficientNetB0..... 61
5.9.2 ResNet50..... 61
5.9.3 InceptionV3..... 61
5.9.4 Xception..... 62
5.10 Model Training Strategy..... 62
5.11 Hyperparameter Configuration..... 63
6. Experimental Setup..... 65
6.1 Hardware and Software Environment..... 65
6.2 Experimental Configuration..... 66
6.3 Training Protocol..... 67
6.3.1 Optimizer and Learning Rate..... 67
6.3.2 Batch Size and Epochs..... 68
6.3.3 Early Stopping..... 69
6.4 Model Parameters and Computational Complexity..... 69
6.4.1 Number of Parameters..... 69
6.4.2 Computational Efficiency..... 70
6.5 Evaluation Procedure..... 71
7. Results and Performance Analysis..... 72
7.1 Overall Performance..... 72
7.2 Model Performance Comparison..... 73

7.2.1 Accuracy Comparison....74
7.2.2 Precision, Recall and F_1-Score Comparison....75
7.2.3 Loss Comparison....76
7.3 Class-Wise Performance Analysis....77
7.3.1 Very Bad and Bad Classes....78
7.3.2 Average Class....79
7.3.3 Good and Excellent Classes....79
7.4 Confusion Matrix Analysis....81
7.5 Training and Validation Analysis....83
7.5.1 Accuracy Curves....83
7.5.2 Loss Curves....84
7.6 Error Analysis....86
7.6.1 Errors Between Adjacent Quality Classes....86
7.7 Computational Efficiency....87
7.8 Comparison with Existing Research....88
8. Discussion....91
8.1 Interpretation of Results....91
8.2 Why the Proposed Multi-Resolution Model Performs Better....92
8.2.1 Contribution of Fine-Grained Features....92
8.2.2 Contribution of Mid-Level Features....93
8.2.3 Contribution of Global Features....93
8.3 Analysis of Borderline Quality Classes....93
8.4 Comparison with Transfer Learning Models....94
8.5 Practical Applications....95
8.5.1 Educational Feedback Systems....95
8.5.2 Automated Grading and Learning Tools....95
8.5.3 Document and Forensic Analysis....96
8.6 Strengths of the Proposed Framework....96
8.7 Limitations of the Proposed Framework....97
9. Conclusion and Future Work....99
9.1 Conclusion....99
9.2 Summary of Findings....100
9.3 Research Contributions....100
9.4 Limitations....101
9.5 Future Research Directions....102
References....103

List Of Tables

Table 1: Comparative Analysis of Previous Studies..... 27
Table 2: Annotation and Quality Assessment..... 47
Table 3: Validation and Test Sets..... 49
Table 4: Dataset Splitting Strategy..... 55
Table 5: Hyperparameter Configuration..... 63
Table 6: Hardware and Software Environment..... 66
Table 7: Experimental Training Configuration..... 68
Table 8: Overall Performance..... 72
Table 9: Model Performance Comparison..... 74
Table 10: Class-Wise Performance Analysis..... 78
Table 11: Errors Between Adjacent Quality Classes..... 86
Table 12: Computational efficiency..... 87
Table 13: Comparison with Existing Research..... 88

List Of Figures

Figure 1: Confusion Matrix of the proposed model..... 41
Figure 2: Data collection process..... 44
Figure 3: Data augmentation..... 51
Figure 4: Research Framework..... 54
Figure 5: Dataset description and preprocessing..... 54
Figure 6: Proposed Multi-Resolution CNN Architecture..... 56
Figure 7: Overall performance of the proposed model..... 73
Figure 8: Accuracy comparison across all the models..... 75
Figure 9: Precision, Recall, and F_1-Score comparison..... 76
Figure 10: Loss comparison across all models..... 77
Figure 11: Per-class metrics of the proposed model..... 79
Figure 12: Class-wise performance heatmap..... 80
Figure 13: Confusion Matrix of the proposed model..... 81
Figure 15: Training & Validation accuracy curves..... 83
Figure 16: Training & Validation loss curves..... 84
Figure 17: Training curves comparison across all models..... 85
Figure 18: Comparison with existing research of Bengali handwriting..... 89
Figure 19: Proposed model's performance summary..... 90

Abstract

This paper presents a multi-resolution feature integrated framework using Convolutional Neural Networks for automated Bengali handwriting quality assessment. The complex characters of Bengali handwriting poses a unique challenge, the horizontal matra line which connects the different characters, compound character formations and substantial inter-writer variability. While a vast amount of research is available to address Bengali character recognition, handwriting quality still remains vastly unexplored. This framework integrates visual features extracted at multiple spatial resolutions through three parallel convolutional streams: a fine grained stream which captures stroke level details, a mid level stream which analyzes curvature and junction patterns and a coarse stream which encodes global shape and spatial layout. These three streams are then concatenated and split into five quality classes: Very Bad, Bad, Average, Good and Excellent. Evaluated on a curated dataset of images of Bengali handwritten characters, the proposed architecture attained an accuracy of 96% with a precision of 95.8% which outperformed the likes of EfficientNetB0, ResNet50, InceptionV3, and Xception baselines by margins of 3.9 to 10.8 percentage points while maintaining only 8.2 million parameters. The results demonstrate that domain specific multi-resolution architecture offer advantages over generic models for handwriting quality assessment, establishing a foundation for automated educational feedback systems and extending applicability to similar complex South Asian scripts.

1. Introduction

Bengali handwriting analysis is a significant field of computer vision due to the existence of the large variability in stroke formation, shape, size, orientation and writing style in handwritten characters. While there are extensive studies that have attained high rate of digits, characters and words recognition in Bengali, comparatively less attention is paid to the recognition of the written character. The quality of handwriting is a problem to be solved, and it involves visual aspects of the writing like stroke regularity, curvature, structural correctness, alignment and general consistency. In this study, the authors tackle this issue using a deep learning framework developed exclusively for Bengali handwriting quality assessment. A multi-resolution CNN is suggested for extracting complementary information from fine, intermediate and global visual aspects and classify handwriting into five quality classes: Very Bad, Bad, Average, Good and Excellent.

1.1 Background

In the field of education, administration, banking, legal documentation and archival work handwriting remains a significant part of the education process. The automatic handwriting analysis is, therefore, an important research field in computer vision, pattern recognition and optical character recognition, and it is still an important field of handwritten information that has to be stored, analysed and converted to a digital form. In recent years, deep learning approaches, particularly Convolutional Neural Networks (CNNs) have made great strides in the machine learning capacity to learn valuable visual features directly from handwritten images [2].

Handwritten Bengali characters pose a special challenge due to the variation in shape, size, stroke thickness, orientation and style of writing in the handwriting. Different writers of the same character can produce very different looking characters that still have the same meaning linguistically. Specialized CNN architectures have thus been developed for the Bengali handwritten characters, digits, modifiers and compound characters [3]. Later models have also emphasized the recognition of more complex Bengali character structures with deeper and more task-specific architectures [5].

Existing research on Bengali handwriting has mostly focused on the problem of recognition, which involves extracting the intent behind the handwriting, namely which digit, character, grapheme or word is written. Unconstrained Bengali handwriting recognition has also been explored using the sequence based techniques for dealing with naturally written script with large structural variation [7]. But identifying a character isn't always a sign of clarity or correctness in writing the character.

Handwriting quality assessment is a different problem. It doesn't look at the content itself, but at the characteristics of its formation, consistency, size, spacing, alignment, and the overall quality of the visuals. Different levels of handwriting quality have already been distinguished using structural handwriting such as character size, slant and spacing [10]. Therefore automated Bengali handwriting quality assessment is more relevant in the field of educational feedback, handwriting improvement system, and automation evaluation system.

1.2 Bengali Handwriting and Script Characteristics

The basic characters, modifiers, compound characters, curves, junctions and connected characters of Bengali handwriting are very rich. These attributes give rise to a much greater visual difference than the simpler isolated character writing systems. Variations in handwriting styles also vary the proportions, curve direction, stroke placement, and general shape of the same character [5].

The 2 dimensionality is another important feature of Bengali writing. Sometimes graphical elements don't go in a left to right order. Some parts of the character may be placed above, below, before or after the main body of the character depending on the character formation [7]. Accordingly, a Bengali handwritten character's visual structure is not only based on the individual strokes but has also taken into account the spatial relationship between the various components of the character.
This complexity is further accentuated by the Bengali written grapheme structure where one written character might be composed of one or more of the consonant grapheme elements required to represent the consonant root, and one or more of the vowel grapheme elements required to represent the vowel root. These combinations are very different in handwritten forms and make recognition and evaluation more difficult. The system should take into account the completeness of the individual components and the uniformity of their spatial distribution in order to evaluate the quality.

1.2.1 Basic Characters, Modifiers and Compound Characters

Bengali script is made up of simple vowels and consonants, vowel modifiers and a huge population of compound characters. The number of visual patterns that can be analyzed by an automated system is substantially higher with these various categories of symbols [7]. The foundation of Bengali writing is basic characters and modifiers are used to modify the visual structure of a character by adding or positioning itself differently around the basic character.

Modifiers can be placed on the left side, right side, top or bottom of the main character and greatly alter the look of the full written unit [17]. Their location, proportion and relationship to

the overall handwriting style thus influence the quality of handwriting. Even if the modifier is wrong, if it is poorly aligned or disproportionately formed, the character will not be as visually pleasing.

Compound characters present an extra level of complexity as they are made from a combination of character components in a more complex structure. They can have multiple curves, junctions, overlapping parts or compressed character parts in their shapes. A compound Bengali character is generally more complex in terms of its structural variation and similarity among its forms than a simple basic character, thus making the recognition of compound Bengali characters harder than that of simple basic Bengali characters [5].

These variations are crucial when evaluating the quality of handwriting, since the quality of a compound character can be determined by more than just the final recognizable shape that it ends up with. The relationship of its components, the continuity of the strokes, proportions, and the formation of the junctions also affect the clarity and good formation of the writing. As a result, both local information and the entire character structure should be taken into account when assessing.

1.2.2 Matra, Curvature and Structural Complexity

The matra is one of the most unique visual features of Bengali writing that is also called a "horizontal headline". It is a crucial element in the top portion of many Bengali characters and can link several characters in handwritten words [7]. In natural handwriting though, the matra can be slightly curved, tilted, broken, thick or uneven depending on the writer.

The matra can vary so much that it can have a direct impact on the appearance of handwriting. Well-formed matra is usually a sign of a more organized character structure while a poorly connected matra or an irregular matra may cause the handwriting to look inconsistent. There is variation in hand writing and the automated system should be able to differentiate between variations that are acceptable and those that are significant when it comes to structure.

Bengali characters have also many curved strokes and loops, intersections and direction changes. These properties make curvature an important visual property to differentiate various handwritten forms [16]. When a character's identity is still discernible despite some curvature and/or stroke direction change, small changes in how that character is curved and/or how the strokes are directed can produce a different look.

The handwritten structure of Bengali is thus complex on multiple levels. Fine structure: thickness of the stroke, the smoothness of its edges and fine irregularities; intermediate structures: curves, junctions and relationships among the components being created; general structures: overall shape, proportion, orientation and spatial arrangement. Such a mix of local and global features

makes multi-resolution feature extraction suitable for handwriting quality evaluation of Bengali, in which the features of various visual scales can be combined to get a more reliable handwriting quality assessment.

1.3 Handwriting Quality Assessment

Handwriting quality assessment aims at the assessment of the quality of handwriting and not the text itself. A recognition system can recognize a character even if it is not well written, and a quality-assessment system must judge whether the pictorial structure is clear, consistent, and acceptable. Structural consistency, slant, character size and spacing can be successfully measured to quantify handwriting quality [10].

Good handwriting is reasonably consistent in shape, size, direction and spacing. These properties can change if writing is not of a high standard even if it is readable. Character uniformity in terms of size and slant along with spacing uniformity can thus give helpful information in automatic character quality assessment [10].

Handwriting quality assessment is particularly beneficial in educational applications. If students receive feedback on poorly formed writing in a timely fashion and gradually practice their writing skills, they can benefit from these [10]. Automated assessment can also enable this feedback to be more consistent and reduce manual assessment of continuous evaluation.

When it comes to Bengali handwriting, more focus should be paid to the stroke shape, curvature, character proportions, consistency of matra formation, and the spatial layout of the various components of the script. These attributes must be assessed as a whole since a character may be identifiable even with multiple structural flaws.

1.4 Challenges of Bengali Handwriting Quality Assessment

Due to the natural variation of handwriting and the complex character structure, Bengali handwriting quality assessment is challenging. In Bengali, there are simple characters, modifiers and also a huge number of compound characters, many of the latter have complicated shapes and cursive characters.

This is a harder problem in handwritten data, because the same character might look different from different hand-writers. There can be differences in size, shape, orientation, curvature, stroke thickness, placement of the components and overall proportion [16]. Many of these differences are not necessarily indicators of poor handwriting, but normal writing styles. A quality assessment system has to thus differentiate between acceptable variation and actual structural degradation.

Another challenge is that the quality of handwriting is not always well-defined. Samples at the extremes of categories (Bad and Average or Good and Average) may have only slight color variations. So small local defects and overall structure of the character must be analyzed to do the effective assessment.

1.4.1 Inter-Writer and Intra-Writer Variability.Inter-Writer and Intra-Writer Variability

Inter-writer variability is a variation in handwriting among different writers. Each person will have his own writing style, with variations in character size, direction of strokes, stroke curves, slant and formation of the components. The characters of Bengali handwritten can thus be clearly distinguished in terms of their appearance yet they represent exactly the same character [16].

Other patterns in handwriting are also strong enough to be used for identifying or verifying the individual writers [18]. This shows the effect of personal writing styles on the visual characteristics of handwritten Bengali works.

Intra-writer variability is when the same writer produces a different version of the same character at different times. Various factors, such as writing conditions, rate, pen movement, attention, and other factors, can lead to variations within the writing of a person [18]. Therefore, it is not possible to establish a standard layout for a well-written sample and a quality-assessment model should not demand this.
This variation is an important problem for automated assessment. The system should be tolerant of the individual writing style while identifying some irregularities such as strokes, distortion, poor proportions, and more, which actually lowers the quality of the handwriting.

1.4.2 Similar Character Shapes and Structural Ambiguity

There are a few Bengali characters with similar visual structure and some characters may be written more than once in different acceptable ways. There are more basic and compound characters, which makes the possibility of similarity of visual form between different classes higher [2].

This is a more challenging problem if the characters that are being composed change in part or all after being combined. In some cases, compound forms can be so complicated that the original consonant parts are not obvious to see [5].

Ambiguities also exist in the case of continuous handwritten characters, where adjacent characters in the Bengali script can overlap and there is no distinct separation between the

characters [7]. These traits can cause confusion in determining if it is a bad handwriting or a true style of writing or if it is a combination of multiple characters.

Small structural differences are significant in quality assessment. If a curve, junction, modifier or matra is slightly distorted, the resulting character may not be of the same quality, but it will not be unrecognizable. The model must, therefore, reflect nuances, rather than just general character forms.

Subjectivity of Quality Evaluation Part of handwriting quality is perceptual and thus the assessment of handwriting is subjective. Borderline handwriting samples may not be evaluated as the same Quality Level by two evaluators. However, there are measurable structural features that can represent perceptual handwriting quality [10].

Some of the properties of handwriting (consistency of slant, size of character, spacing, uniformity of structure) give more objective information to evaluate handwriting [10]. These measurable properties can be modified to suit Bengali characters (stroke regularity, shape correctness, curvature, alignment and consistency of structure).

When multiple quality levels are employed rather than just good versus bad decision, subjectivity is particularly significant. The distinction between the Very Bad and Excellent handwriting is likely to be more clear, whereas the Average to Good handwriting distinction can be a lot more subtle. Thus, there is a need for consistent criteria for annotation, to minimise disagreements and ensure training labels are reliable.

1.5 Motivation of the Study

Bengali digit, isolated character, compound character and handwritten word recognition have made significant progress in the field of Bengali handwriting recognition. CNN based lightweight architectures have been proved to have good performance in Bengali handwritten character recognition [2]. Even more complex architectures have dealt with basic characters, modifiers, numerals, and complex compound characters [5]. Recurrent neural network for continuous handwritten text has been further explored in case of unconstrained Bengali handwriting [7].

Notwithstanding the developments, the majority of literature in Bengali reviewed is focused toward the question of what has been written. Less emphasis has been placed on determining if the handwriting itself is well formed. Previous general research in handwriting has demonstrated that a structure can be used to give meaningful handwriting quality indicators [10]. The importance of extending this concept to Bengali is that it includes challenges of curves, compound structure, modifiers, matra, and there is a substantial writer variation.

Thus, the primary aim of this study is to create an automatic framework that is able to assess Bengali handwritten characters based on their visual quality. The proposed approach is not single-featured, but instead takes into account information on various spatial scales. Fine details can indicate irregularities at the stroke level, medium level features can indicate the shape and pattern of the junctions, and wide level features can indicate the character shape and proportion.

Such a system may be useful in the educational field where students need to be instructed on character development. It can also serve as the basis for handwriting-practice and auto-evaluation software for digital handwriting. The primary aim is to go beyond the mere recognition of handwritten words to a handwritten word recognition system that can also assess the quality of the handwritten form.

1.6 Problem Statement

Most of the existing Bengali handwriting recognition systems are able to recognize digits, characters, graphemes or words. These methods can be used to identify handwritten text very accurately, but they don't necessarily tell you if a character has been written well or correctly. A classification model may be able to recognize a poorly formed character successfully.

Handwriting assessment for Bengali language is even more complex due to the significant differences in character shape, size and style of handwriting of every person [16]. There is also a lot of compound characters and similar characters with similar shapes, which increases the structural complexity [2]. Further, the visual structure can be more complicated when dealing with natural handwritten samples, which may include elements that overlap or are misaligned [7].

Traditional handwriting quality methods that rely primarily on structural properties like slant, spacing, and character size show that the quality of handwriting can be measured [10]. But Bengali has multiple levels of visual information such as the fine strokes, curvatures, junctions, and the complete structure of the characters. Relying on just one feature scale would thus not be able to capture all of the quality aspects of Bengali handwriting.

The problem of this study is to develop an automated approach to classify Bengali handwritten characters of five different quality levels Very Bad, Bad, Average, Good and Excellent. The system should be able to detect the fine stroke irregularities, intermediate structural patterns, and global shape characteristics with robust performance to natural variations in individual handwriting. The complementary visual features are therefore integrated using a multi-resolution CNN framework and consistent Bengali handwriting quality assessment is provided.

1.7 Research Objectives

The main goal of this work is to create an automatic deep learning approach to evaluate the quality of handwritten Bengali characters. The study is intended to study various visual traits of handwriting and to classify handwriting samples into various levels of quality, by analyzing the structural and visual features of each sample.

The main difference between this research work and conventional handwriting recognition system of Bengali is that the latter is only for the recognition of written character while this research is about the evaluation of the quality of the character formation. The proposed approach is to extract information from multiple levels of space, such as fine stroke information, structural pattern information, and overall appearance of the character.

1.7.1 General Objective

The overall aim of this work is to design a multi-resolution Convolutional Neural Network (CNN) automatic quality assessment system for Bengali handwritten characters.

Proposed approach aims to categorize Bengali handwritten samples into five quality categories of handwriting namely Very Bad, Bad, Average, Good and Excellent by learning meaningful visual features from handwriting images.

1.7.2 Specific Objectives

The specific objectives of this research are:

1. To develop a Bengali handwriting quality assessment dataset containing samples categorized according to different handwriting quality levels.
2. To analyze important visual characteristics of Bengali handwriting, including stroke patterns, curvature, structural arrangement, and overall character shape.
3. To design a multi-resolution CNN architecture capable of extracting handwriting features at different spatial levels.
4. To develop separate feature extraction streams for:
    - Fine-grained stroke-level information,
    - Mid-level structural features such as curves and junctions,
    - Coarse-level global character structure.
5. To combine multi-scale features through an effective feature fusion strategy for improved quality classification.
6. To compare the performance of the proposed framework with existing deep learning architectures, including EfficientNetB0, ResNet50, InceptionV3, and Xception.

7. To evaluate the proposed model using standard performance metrics such as accuracy, precision, recall, and F_1-*Score*.
8. To analyze classification errors and identify challenges associated with different handwriting quality categories.

1.8 Research Contributions

The major contributions of this research are summarized as follows:

1. **Development of a Bengali handwriting quality assessment framework:**
   This study introduces an automated approach for evaluating Bengali handwriting quality, addressing a research area that is different from conventional Bengali handwriting recognition tasks.
2. **Multi-resolution feature learning approach:**
   A multi-resolution CNN architecture is proposed to capture handwriting characteristics from multiple levels. The framework analyzes fine stroke details, intermediate structural patterns, and global character shapes simultaneously.
3. **Five-level handwriting quality classification:**
   The proposed system classifies Bengali handwritten samples into five quality categories: Very Bad, Bad, Average, Good, and Excellent, allowing more detailed assessment compared with simple binary classification.
4. **Integration of complementary handwriting features:**
   The proposed approach combines different types of visual information through feature fusion, enabling the model to better understand complex Bengali character structures.
5. **Performance evaluation against existing deep learning models:**
   The effectiveness of the proposed framework is evaluated by comparing it with established architectures such as EfficientNetB0, ResNet50, InceptionV3, and Xception.
6. **Foundation for automated Bengali handwriting feedback systems:**
   The proposed framework provides a foundation for future educational tools and digital platforms that can automatically evaluate and provide feedback on Bengali handwriting quality.

1.9 Scope and Limitations

There are some key areas of scope to be discussed so are the limitations.

1.9.1 Scope of the Study

In this research, the automatic evaluation of the quality of handwritten Bengali characters is done by applying deep learning technique. The key areas of scope are:

Developing Computer vision Based framework of Bengali Handwriting quality classification.

Evaluation of visual writing attributes (stroke formation, curvature, consistency of structure, shape of characters).

Handwritten samples are classified into five pre-defined quality categories: Very Bad, Bad, Average, Good, and Excellent.

Application of CNN based feature extraction methods for learning Handwriting Representation.

Investigation of multi-resolution feature learning for capturing information from different spatial levels.

Evaluation of the proposed approach with comparison to the existing models of Transfer Learning.

The study concentrates mainly on isolated Bengali handwritten characters and not on handwritten sentences or handwriting evaluation in a document level. The proposed framework is meant as a benchmark for upcoming handwriting assessment systems for educational and digital learning settings.

1.9.2 Limitations of the Study

The proposed model is effective for assessing the handwriting quality of Bengali characters but some limitations are found:

Dataset size limitation: Deep learning models rely heavily on the amount and variety of data used to train the model. The study is a limited one, as a larger data set of writers and writing conditions could enhance the generalizability.

Limited handwriting categories: This study has been conducted on five levels of quality. Handwriting quality may have one or more finer levels of judgement and be continually assessed as a human perception rather than in fixed categories.

Character-level assessment only: The existing system is based on the assessment of isolated handwritten characters. The handwriting evaluation is at word, sentence, and document levels, which are not included in this research.

Writer diversity: There may be further variations not fully representative of these differences in age group, educational background, writing tools or writing environment.

Suspicious of images: The performance of handwriting feature extraction and model may be influenced by variations in the scanning conditions, image resolution, lighting and noise.

No real-time deployment evaluation: Therefore, the aim of the current research is to develop and analyze the model. For mobile or web-based educational applications, further optimization and testing is required for practical deployment.

The constraints outlined above offer scope to researchers to create more general and comprehensive handwriting quality assessment systems for Bengali.

2. Literature Review

The progress of machine learning, computer vision, and deep learning have significantly improved handwriting analysis for the Bengali language. The research has advanced from the handwritten digit recognition to simple and composite character classification, word-level OCR, unconstrained handwritten recognition, grapheme based modelling and handwritten writer oriented analysis. In this section these developments are reviewed systematically, and the focus is on datasets, recognition architectures, sequence-based methods and handwriting quality assessment, limitations of existing methods and the remaining research gap.

2.1 Overview of Bengali Handwriting Research

In the early periods of research in Bengali handwriting recognition, the focus was primarily on handwriting recognition of isolated digits, as it has a comparatively small number of classes and yet has a great variation in writing styles. In a comparative deep CNN study, AlexNet, MobileNet, GoogLeNet/Inception V3 and CapsuleNet were compared over the database NumtaDB and it was shown that deep learning could achieve very high recognition performance even on a large dataset of augmented Bengali numerals [1].

Research then went beyond just the digits and included other Bengali symbols. Low cost Bengali numeral, modifier, basic and compound characters architecture, BengaliNet developed. The network utilized about 2.24–2.43 million parameters and it was assessed on different formations of the CMATERdb as well as tested on Ekush, BanglaLekha and NumtaDB [2].

EkushNet was another step towards generalized Bengali character recognition. The model did not focus on any single category, but rather on basic characters, numerals, modifiers and compound characters. Another important aspect of Bengali handwriting research underlined by the use of Ekush and CMATERdb was the need for large standardised datasets [3].

The survey spans 20 years and a comprehensive study of the handwritten digit recognition of Bengali was conducted, covering dataset characteristics, digit ambiguities, techniques in preprocessing, feature extraction, machine learning, CNN, transfer learning and real-life applications. The survey revealed that the recognition accuracy had significantly improved but these were the challenges that were still significant, namely variation in writing style, incomplete strokes, redundant strokes, noise and differences in the datasets [4].

Then, recognition research became more complex. RATNet presented a residual learning method and spatial attention mechanism for numeral, basic character, modifier and compound character recognition on various datasets in Bengali. It demonstrated that numerical and basic-character recognition have reached very high accuracy, but compound characters are relatively hard to recognize [5].

Further expansion of class space was achieved with the BNVGLENET where vowels, consonants, numerals, compounds and kar-fola forms were taken into account. A custom dataset for handwritten bengali character recognition was prepared which contained hundreds of differently shaped classes, thus proving that the recognition of Bengali handwritten characters is a practical problem and this should eventually be extended from the limited number of classes that are used in other previous studies [6].

The study also started to shift from symbolic studies to studies in the wider context. An unconstrained Bengali handwriting recognition framework used BLSTM-CTC to recognize the handwritten text lines and incorporated semi-ortho-syllable units to represent the complicated Bengali writing system in a more effective manner [7]. The lower recognition rate achieved for continuous text as compared to isolated character recognition showed the higher difficulty of the segmentation-free sequence recognition.

This transition was facilitated by the creation of the datasets. In addition to pre-segmented handwritten characters, CMATERdb1 also has ground truth line segmentation for handwritten Bengali and Bengali-English mixed script document pages, which enabled the investigation of handwritten documents instead of only handwritten characters [8].

The end-to-end handwritten Bengali word recognition task was then studied using the combination of CNN feature extractors with recurrent layers and CTC loss. This helped to overcome the need of explicitly segmenting the characters in the Bengali handwritten images and proved the capability of sequence-learning based OCR techniques in Bengali OCR [9].

Another area of research focused on handwriting (not handwriting recognition). Good versus bad handwriting is distinguished using structural properties, such as slant, character spacing, and character size, in machine-learning classifiers. This study was performed on English handwriting, but it provided measurable structural features which could be relevant to handwriting-quality analysis in other scripts [10].

The lightweight CNN development was also continued. BornoNet proved the effectiveness of a simple CNN for recognition on the CMATERdb, ISI, BanglaLekha-Isolated and mixed datasets [11].

Research on handwriting was also continued in the direction of writer analysis. In this case, the textural characteristics derived from modified FFT, GLCM and DCT features were investigated for Bengali writer verification by taking handwritten image from 100 writers [12].

A large-scale comparison of the performance of 11 state-of-the-art CNN architectures was then performed, showing that the performance of the models is significantly different depending on network architecture and recognition task. Basic characters, digits, modifiers and compound characters from the large Ekush dataset [13] were used in the experiments.

Another script has been also investigated as handwriting recognition using object detection. A sequential character detection based YOLOv3 handwritten word recognition system was found to be a possible alternative to the common word recognition without using a lexicon [14]. This is not a Bengali language specific work, but gives a relevant methodical approach to future work on segmentation free recognition of complex scripts.

In recent Bengali research, computational efficiency has been still stressed. A small custom CNN with only around 0.8 million parameters achieved good performance, on a small dataset of handwritten characters, with CMATERdb, which means that Bengali handwritten recognition may be implemented with the help of comparatively low cost computational resources [15].

Reconsideration of the representation of Bengali handwriting itself was done through the Bengali.AI grapheme dataset. The data is not just a set of Bengali characters in a sequence, but rather grapheme roots and vowel and consonant diacritics with over 411,000 examples of 1295 common graphemes [16].

Lastly, a PDF-CNN and Siamese network jointly was developed for Bengali writer verification in the field of hand writing analysis. The textural and allographic probability distributions were manually created and combined with CNN features learned from the data, showing that structural handwritten information and CNN features complement each other [17].

A large scale Bengali hand writing database was also created with samples of handwritten characters and aesthetic-quality assessments to show that Bengali hand writing databases can be used for quality-oriented analysis as well as hand writing recognition [21]. The aesthetic quality of handwritten characters has been shown to be measurable by global features such as shape and component-layout [22].

Generally, the literature shows that recognition of isolated textbooks is very successful, while the recognition of handwritten documents, script-aware analysis and handwriting-quality assessment are slower.

2.2 Bengali Digit and Character Recognition

The fields of digit recognition and isolated-character recognition are the most extensively studied fields in Bengali handwriting recognition. It can be said that the tasks usually assume that each symbol has already been segmented from the original document, so that the recognition system can focus on classification. This makes the OCR problem easier, but this cannot be neglected because there is still significant variation due to different writing styles, poorly defined shapes, stroke differences and complexity of compound characters.

2.2.1 Handwritten Digit Recognition

A comparative study of AlexNet, MobileNet, GoogLeNet/Inception V3 and CapsuleNet has exhibited the effectiveness of using deep CNN for Bengali digit recognition. AlexNet outperformed the other reported architectures in terms of accuracy (99.01%), and was also the fastest in terms of compute time (1.14 seconds), using NumtaDB. MobileNet, GoogLeNet and CapsuleNet took 12.52, 22.53 and 3.86 seconds respectively [1].

BengaliNet was then found to perform well on various digit sets. It achieved an accuracy of 99.01%, 99.13%, 98.11% and 97.50% on the ten class CMATERdb numeral database, Ekush numerals, BanglaLekha numerals and NumtaDB [2] respectively. The difference between the datasets suggests that the model architecture is not the only factor which influences reported recognition performance, but also the complexity of the datasets and their sample characteristics.

EkushNet also introduced Bengali numerals to a large scale 122 class recognition problem with basic characters, compound characters, modifiers and digits. The total number of digit samples included in the Ekush dataset used in this work was around 30,830, which shows the shift from digit specific databases to larger multi class Bengali recognition databases [3].

In a survey of handwritten Bengali digit recognition, it was found that there are several sources of confusion for digit recognition from a morphological point of view. One numeral may look like another numeral due to similar baseline structures, incomplete strokes and unnecessary or extended strokes. Additionally, the survey highlighted that there are significant differences in the difficulty of data, the number of writers, noise, and preprocessing of datasets like CMATERdb, Ekush, BanglaLekha-Isolated, and NumtaDB [4].

The numeral-recognition performance was especially good for RATNet. It was reported that the methods of residual and spatial-attention mechanism can achieve an accuracy of 99.66% on the numerals in the CMATERdb and 99.80% on the numerals in ISI [5]. The findings confirm the statement that recognition of isolated Bengali numerals is at very high level when tested under a controlled environment of a benchmark.

The expanded BNVGLENET was also extended to cover a wide handwritten character space including Bengali numbers. The numeral category with 99.1% accuracy and an F_1-*Score* of 99.3% was reported in its category-level evaluation [6].

The digit part of Ekush was also employed in a more extensive character study in a later large scale assessment of contemporary CNN designs. The results showed that the accuracy and computational aspects of different deep architectures vary, which highlights the need for architecture selection and not to think that all deep CNNs are alike [13]

Finally, a small customized CNN showed that one can also solve the problem of recognition from the standpoint of cost of computation. The architecture had around 0.8 million parameters indicating that smaller networks are still a viable path in deployment environments where constraints matter [15].

Overall, these studies indicate that the problem of digit recognition from bangla handwritten isolated digit is not just a high recognition rate problem. There is a need to pay more attention to cross-dataset robustness, noisy writing, conditions of real documents, and efficient deployment.

2.2.2 Isolated Character Recognition

BengaliNet extended the recognition to 50 basic characters, modifiers and some combinations of compound characters. It's 98.97% accurate on CMATERdb for 50 basic characters, and 98.35% accurate on one 171-class compound-character configuration. For more challenging expanded configurations, performance was reduced to 95.89% and 95.71% respectively indicating that recognition became increasingly difficult with the increase in number and diversity of the classes [2].

EkushNet covered 122 classes of characters, numbers, modifiers and compound characters. It was found that the input images were normalized to a size of 28 × 28 and the network used convolution, max pooling, batch normalization, dropout, and fully connected layers. It has attained 97.73% accuracy on Ekush and 95.01% cross-validation accuracy on the CMATERdb [3].

The survey also covered Bengali digit recognition, where features were learned from natural images instead of manually designed feature pipelines and CNN was replacing them with the passage of time [4].

In RATNet, different categories of characters were explicitly examined and there was a clear relationship between structural complexity and recognition difficulty. It performed well on CMATERdb with basic characters, modifiers, and compound characters accuracy rates of 99.27%, 98.78%, and 97.70%, respectively. When it was applied to a newly collected compound-character set, the accuracy decreased slightly to around 93.42%, which showed that compound characters are still very challenging [5].

The recognition space was significantly expanded by BNVGLENET. It used a specially created set of images containing vowels, consonants, compounds, numerals, kar-fola forms, and hundreds of different classes. The proposed model achieved an accuracy rate of 98.2% in the grouped vowel-consonant task and an accuracy rate of around 97.5% in the large individual-class task [6].

In addition to this, Boronet offered a light solution for the 50 basic characters of Bengali language. The best results obtained by it on the CMATERdb was 98%, within ISI 96.81%, BanglaLekha-Isolated 95.71%, and mixed dataset [11] 96.40%.

Later, eleven CNN architectures were studied, and basic characters and the full Ekush dataset were assessed. For the basic-character experiment, ResNet152V2 achieved recognition accuracy of 96.68% while NASNetMobile achieved the lowest recognition accuracy of about 87.97%, showing that the complexity of the architecture and design of the network has significant impact on the recognition of Bengali character [13].

Additionally, the performance of the relatively small network further illustrated that useful performance could be obtained from a cost-effective custom CNN. The model achieved 95.67% accuracy using 3,000 custom character samples and 93.83% accuracy using evaluation on the CMATERdb 3.1.2.The model achieved 95.67% by using the 3,000 custom character samples and 93.83% accuracy by evaluation on the CMATERdb 3.1.2.

Finally, the Bengali.AI work was able to question the idea that the smallest possible recognition unit should always be a Unicode character. Bengali graphemes may have roots as well as vowel and consonant diacritics, but the visual position of the diacritics is not necessarily linear with their encoded position. Thus, representations of such combinations as graphemes gives a script-aware perspective to recognition [16].

The results indicate that basic character recognition is quite well developed, but scaling up to full Bengali writing system is significantly more challenging.

2.3 Bengali Word and Text Recognition

The recognition of complete Bengali words and handwritten text is a basic problem much more complex than handwritten character classification. This includes variable length sequences, adjacent symbols, modifiers, compound symbols, matra relationships, writing slant and ambiguous boundaries between co-adjacent symbols.

2.3.1 Word-Level Recognition

Studies on free handwriting in Bengali language revealed that the holistic word recognition is hard due to the huge vocabulary and morphological variation of the language. It was therefore proposed that using memorisation of whole words would need massive amounts of training data [7].

The creation of CMATERdb1 was important for word- and document-level research. It has handwritten pages, not only handwritten characters, but natural line and word structures. The

database also contains Bangla-English mixed documents, which adds to the recognition challenges due to structural differences between Bangla and English [8].

The subsequent step was an end-to-end handwritten word-recognition system in Bengali that used CNN to extract features from handwritten words paired with recurrent sequence modeling to mitigate the need for explicit character segmentation. For visual feature extractors, DenseNet, Xception, NASNet, and MobileNet were tested and for sequence modeling, LSTM and GRU were explored. The best model DenseNet121-GRU performance is obtained in BanglaWriting dataset with character error rate of 0.091 and word error rate of 0.273.

YOLOv3 was tested for handwritten English words in a related work without the use of a lexicon. Recognized characters instead of the whole word as a single vocabulary item was done by using object detection [14]. This approach was not tested on Bengali, but the idea is still useful because the idea of decreasing the use of pre-established lexicons may be beneficial for languages with extensive vocabulary and complex word structures.A privacy-preserving cultural-linguistic authentication framework for autonomous delivery in low-resource environments was introduced by Mondal et al., addressing both security and accessibility challenges in under-connected regions [31]. Recent Bengali handwritten word recognition work has considered inter-person and inter-age variation in words recognition using Swin Transformer, BiLSTM and CTC model, which has achieved significantly better recognition performance on student handwritings [35].

The current evidence thus indicates that word recognition in Bengali is still far behind in maturity compared to the isolated character recognition, especially without explicitly assuming regular segmentation and a fixed vocabulary.

2.3.2 Unconstrained Handwriting Recognition

Using a BLSTM-CTC approach and a newly introduced semi-ortho-syllable representation, a significant progress has been achieved in unconstrained Bengali handwriting recognition. The system handled whole text-line images, and not previously-segmented characters. Preprocessing steps performed on the image are binarization, deslânting, normalization of height and feature extraction of the HOG in a sliding window. The recurrent network consisted of two BLSTM layers with around 200 neurons, and a CTC layer with 917 output neurons [7].

The data set consisted of 2338 unconstrained Bengali text lines with about 21000 words and handwriting of over 100 writers. The character-level recognition accuracy of 75.40% was obtained with five-fold cross validation. Substitution resulted in 18.91% of the total errors, deletion in 4.69% of the total errors, and insertion made up about 0.98% of the total errors [7].

Another important aspect of unconstrained recognition, document segmentation, is tackled in the context of CMATERdb1. It comprises of 150 handwritten pages of documents, of which 100 are

written entirely in Bengali language and 50 are written in both Bengali and English languages, with ground-truth data for line segmentation [8]. Bengali words are segmented differently than words in Latin script, because of their structure—especially that of the matra and vertically placed modifiers.

Later, end-to-end word recognition also revealed that the combination of recurrent models and CTC could also eliminate the requirement for symbol segmentation, but the remaining word error rate indicates that the task of natural Bengali handwritten word recognition remains challenging [9].

The grapheme-based representation will be discussed later, and is one of the directions that might be useful for future unconstrained systems. This representation is a structured representation of the root, vowel diacritic and consonant diacritic as part of a grapheme that is closer to the orthography of Bengali than a simple horizontal sequence of characters [16]. Furthermore, a handwritten Bengali text recognition system based on a resource-efficient handwritten Bengali character recognition system has fused the grapheme-based tokenization with a decoder-only Transformer, further demonstrating the shift towards a script-aware Transformer architecture for Bengali handwriting [36].

A CNN-Transformer framework for Bengali handwritten text recognition also proved that CNN based visual feature extraction with a Transformer based sequence modeling can enhance the recognition accuracy without sacrificing the CNN modeling accuracy [39]. It has been shown that the applicability of modern sequence model extends beyond the traditional CNN-RNN pipeline, as encoder-decoder model has been investigated on handwritten and printed Bengali text for optical text recognition using encoder-decoder framework in the field of low-resource language [40].

2.4 Bengali Handwriting Datasets

Standardization of datasets and its advancement in Bengali handwriting recognition have been a major concern. There are, however, significant differences between datasets in classes, number of writers per class, image quality, the preprocessing steps, the amount of balancing, and whether the samples are symbols or natural handwritten documents.

2.4.1 CMATERdb

For recognition, BengaliNet constructed eight data sets from several subsets of the CMATERdb which included numerals, basic characters, modifiers, compound characters. The number of classes varied between 7, 10 and up to 256 classes (depending on the configuration) [2].

The cross-validation was performed with the aid of EkushNet and the CMATERdb. The version used in the study comprised about 15,000 Bengali alphabet images and 6,000 digit images—at first, they were represented with a resolution of 32 × 32 pixels [3].

The survey offers a detailed description of the CMATERdb 3.1.1 of Bengali numerals. It has 6,000 images - 600 images per each digit class [4].

A subset of the CMATERdb was used with RATNet. It was reported that its statistics of the dataset contain 6,000 numeral samples, about 15,000 basic-character samples, 3,656 modifier samples, and 42,959 compound-character samples with 171 compound classes [5].

CMATERdb1 had incorporated the concept of the dataset into something beyond mere classification. It has 150 handwritten document pages, two subsets containing only Bengali script and with Bengali-English mixed script, and contains ground truth for the evaluation of the text-line extraction [8].

Another main benchmark for Boronet was the use of CMATERdb, and they claim to have reached 98% accuracy in the dataset [11].

A later low-cost CNN was tested on CMATERdb 3.1.2 with 93.83% recognition accuracy and it showed that the model of CMATERdb has been used as a benchmark for compact CNN as well [15].

Assessing the extent of the use of CMATERdb is beneficial for benchmark testing but note that the use of different versions and experimental splits will affect reported results.

2.4.2 Ekush and BanglaLekha-Isolated

Ekush and BanglaLekha-Isolated were used as two separate sets of data for evaluation. It was able to get 98.36% accuracy for Ekush basic characters, 99.13% accuracy for Ekush numerals, 96.09% accuracy for BanglaLekha basic characters and 98.11% accuracy for BanglaLekha numerals [2].

EkushNet adopted Ekush data as their primary training data set. The study involves the description of about 368,776 pictures, of which 155,570 are samples of the alphabet, 151,607 are compound characters, 30,830 are digit samples, and 30,769 are modifier samples [3].

The later survey introduces Ekush as one of the largest Bengali handwriting datasets that consists of around 367,018 samples of 122 classes. It also includes details of the author, including his or her age, gender, place of birth and educational level. In BanglaLekha-Isolated, there are almost 20,000 numeral samples from around 2,000 writers, and they also include age, gender and an expert-assigned aesthetic-quality index [4]. RATNet reported about 166,105 samples of BanglaLekha-Isolated numerals, basic characters and compound characters; which were used for

cross-dataset evaluation [5]. BornoNet also tested upon BanglaLekha-Isolated and found the accuracy of validation as 95.71% [11].

A comprehensive comparison of 11 CNN architectures followed on the entire Ekush dataset which contained 367,018 images across 122 classes [13]. To overcome the scarcity of Bangla primary school handwriting data, a recently introduced handwritten primary school primary school handwriting (PSPSH) data set was compiled with 24420 isolated character/numeral images collected from 500 students from four districts of Bangladesh [37].

The datasets are of significant importance due to the higher number of writers and the diversity of classes within them, allowing for larger experiments than many previous Bengali handwriting datasets.

2.4.3 NumtaDB and Bengali.AI Grapheme Dataset

A comparison of four deep CNN architectures was carried out in [1] with NumtaDB being the central dataset. It also had a large collection, with an augmentation and a relatively diverse set, which made it a more challenging environment than some of the earlier Bengali numeral sets. In that experiment, AlexNet yielded the best performance [1].

BengaliNet was also independently tested on NumtaDB and its numeral-recognition accuracy was 97.50 % [2].

According to a detailed survey, NumtaDB is a collection of six source databases with over 85,000 images from around 2,700 writers. The training and testing sets were generated in separate partitions with minimal overlap between the sets of contributors, and challenging artifacts like noise, occlusion and rotation were added to the testing set [4].

After that BNVGLENET was tested with an expanded character recognition method using its own created data set and Bengali grapheme resource. The grapheme dataset is compared to the traditional isolate character datasets [6] and found to have over 400,000 images and a larger class space.

The Bengali.A major change in the way of the dataset design is AI grapheme dataset. It has around 411,000 hand-curated samples of 1295 commonly used Bengali graphemes which have approximately 900 uncommon graphemes in the test data for out-of-dictionary evaluation [16]. The graphemes are represented using grapheme roots, vowel diacritics and consonant diacritics, so as to be able to classify them in multiple targets [16].

2.5 CNN and Deep Learning Approaches

Since CNNs are capable of learning visual features from input images, they have been successfully replacing a large number of manually engineered pipelines in the field of Bengali handwriting recognition. However, there are several different approaches that have been reported in the reviewed studies, such as custom lightweight CNNs, deep CNNs, transfer-learning, recurrent sequence models, and hybrid systems based on feature-learning.

2.5.1 Custom CNN Architectures

BengaliNet is specially developed with low cost as an architecture for Bengali handwriting. Its parallel paths employed different convolutional receptive fields and the resulting architecture had around 2.24–2.43 million parameters depending on the number of output classes. The model's performance in all cases was superior to its own convolution paths and a few larger standard CNN architectures [2].

EkushNet employed a tailored multilayer design with convolution, max pooling, batch normalization, dropout and fully connected layers. The convolution paths were merged together in parallel prior to classification and thus the model was able to process all 122 categories represented in Ekush [3].

RATNet added another custom design included residual blocks and spatial attention. The complete network had about 2.89 million parameters which is relatively low as compared to VGG-16 and DenseNet-121 used in the study, and had higher accuracy for several Bengali character categories [5].

BNVGLENET has fused architectural ideas of LeNet-5 model and VGG to tackle a large class room in Bengali. The custom model was specially developed for distinguishing vowels, consonants, compounds, numerals, and kar-fola classes in the same recognition model [6].

Following a light-weight design philosophy, BoroNet exhibited good performance in all four databases namely (CMATERdb, ISI, BanglaLekha-Isolated and a mixed version of the databases [11]).

The later cost-effective CNN decreased the number of parameters to around 0.8 million parameters, with 95.67% accuracy on the custom dataset and 93.83% accuracy on the CMATERdb 3.1.2 [15].

Deep Learning techniques have been used for Bengali compound character recognition, which shows that the CNNs models are effective in recognizing Bengali handwritten compound character [29].

In summary, these studies suggest that task-specific Bengali CNNs can perform competitively with significantly larger generic CNNs.

2.5.2 Transfer Learning Architectures

AlexNet, MobileNet, GoogLeNet/Inception V3, and CapsuleNet models were compared to learn the efficacy of the well-known deep architectures for Bengali digit recognition, but it was mentioned that there is scope for further research on implementation of transfer learning [1].

BengaliNet developed its own model and compared it to some well-known CNN models; and discussed the previous handwriting recognition works done in Bengali using AlexNet, VGG-16, ResNet-50 and other transfer learning techniques. The selected CMATERdb configurations [2] generally yielded better results for BengaliNet, although it had significantly fewer parameters.

CNN-based transfer learning is an important direction highlighted by the Bengali digit-recognition survey and is being used for handwritten Indic/Bengali numeral recognition [4].

Prior to the recurrent sequence processing, DenseNet, Xception, NASNet and MobileNet were studied as CNN feature-extraction back bones for Bengali handwritten words. The best result was obtained by DenseNet121 and GRU [9] together that is DenseNet121.

The comparison of architecture sizes, made directly, was conducted on the architecture DenseNet201, EfficientNetB0, InceptionResNetV2, InceptionV3, MobileNetV2, NASNetMobile, ResNet50, ResNet152V2, VGG-19, Xception and EkushNet [13] that are the largest. The experiments illustrated that high capacity architectures may offer good accuracy but mobile networks might offer more practical recognition performance and computational demands. Multi-model approaches have also been investigated for Bengali handwritten character recognition, which indicated that leveraging multiple learned representations can boost the recognition performance when applied to different Bengali handwritten character datasets [30].

2.5.3 Sequence-Based and Hybrid Approaches to Programming

The literature about digit recognition is showing a growing trend of employing ensembles, hybrid descriptors, and combined CNN-based systems when single-model pipelines are not enough [4].

A more basic innovation was for unconstrained Bengali handwriting using BLSTM-CTC. The one image was not classified, but a series of HOG feature vectors was extracted from a complete line and the corresponding series of characters was generated bidirectionally by CTC without any explicit character-level segmentation [7].

This idea was extended to end-to-end word recognition with a CNN training visual features directly prior to the LSTM or GRU recurrent layers. For the image feature matching to output characters, the best results were obtained by CTC loss (Binarized, DenseNet121-GRU) [9].

Another variant of handwriting analysis classifies hybrid handwriting, but uses handcrafted textural information instead of sequence modelling. Superimposed Bengali character images were used to obtain modified FFT, GLCM and DCT representations to obtain inter-writer and intra-writer variation [12].

This idea was built on with handcrafted probability-distribution maps being transformed using CNNs and then verified using Siamese learning for Bengali writers. The combination of handwriting features with automatically learned deep representations in the textural PDF-CNN-Siamese configuration yielded the lowest EER of 2.36%, confirming the benefits of using these features in tandem [14]. Recently, another Bengali handwriting analysis framework used CNN and BiLSTM models with kinematic, statistical and spatial features, which proved that learned representation and handwriting-specific representation are both useful for writer identification and can be combined to form a hybrid representation for use [38].

2.6 Handwriting Quality Assessment Research

Handwriting quality assessment is still significantly less studied as compared to recognition.

An assessment of the Bengali handwriting datasets suggests one possible useful resource for this task. Besides the writer's age and gender information, BanglaLekha-Isolated has an aesthetic-quality index assigned by several experts. This can be used to give a possible basis to explore hand writing quality outside traditional character labels [4].

Measurable structural properties such as slant, character spacing and word spacing were assessed by a dedicated handwriting quality assessment framework. Both binarized and skeletonized images were used, and statistical representations were extracted. SVM and k-NN classifiers were compared. The highest test accuracy was 98.5% using SVM [10].

The experiment was, however, performed with 322 images of handwritten English text-lines which were scanned, so the features proposed were defined based on the properties of Latin script. The study itself recommended the adoption of handwriting quality assessment in Indian regional scripts (IRS) [10].

The study by Bengali writers and verifier research further confirms that handwritten structure has quantitative information in addition to the textual information. The writer-dependent patterns were captured from a dataset of 500 documents of 100 writers through the modified FFT, GLCM and DCT features [12].

The latest Bengali writer verification has been an amalgamation of textural and allographic probability distribution with CNN and Siamese learning. Best results in terms of EER for the textural PDF-CNN-Siamese configuration are 2.36%, the CNN-MLP with textural data 3.42%, and the CNN-MLP with allographic data 4.07% [17].

The method has also been examined regarding the recognition confidence, where it is found that the confidence obtained from handwriting-recognition systems can be used as information for writing quality [23]. Deep-learning models have been shown to learn quality-related features directly from the handwriting image, instead of using only features that are manually designed, in CNN-based handwriting evaluation [24]. The automatic assessment of handwriting quality by machine learning has also demonstrated that handwriting quality can be measured on the basis of automatically-learned characteristics, and not just traditional handmade measures [25].

Transfer learning-based models have also been used to evaluate handwritten document images, and this shows the potential of deep-learning models for assessing the overall visual quality of handwritten documents [26]. Recent studies of handwriting quality have begun to focus on both structural and perceptual aspects in the same studies and suggest that several complementary properties may yield a more complete assessment of handwriting quality [27]. An active-learning framework has also been considered to study handwriting-quality assessment, focusing on the use of structural, perceptual and marginal features, wherein the importance of using diverse train sets and iterative quality evaluation is emphasized [28].

A handwriting study in Bengali has directly studied both the legibility and aesthetic quality, where human assigned scores are used for supervision of machine-learning models, and both these properties are automatically evaluated from offline Bengali handwriting [32]. In later work, a deep-learning method showed that CNNs are able to estimate human penmanship scores on handwritten character images as a substitute for the subjective handwriting assessment of handwritten character images [33]. Another interpretable deep-learning method with TabNet also showed that aesthetic handwriting evaluation can be based on global shape characteristics and also include other structural elements to enhance prediction and explainability [34].

However, handwriting-quality assessment and writer verification are different problems. Handwriting verification is used to verify whether handwritten samples are from the same person, while handwriting-quality assessment is used to assess handwriting qualities like consistency, readability, spacing, alignment, and formation. The reviewed literature is therefore rich in Bengali recognition, Bengali writer analysis research but lacks Bengali specific research on automatic handwriting-quality assessment.

2.7 Comparative Analysis of Previous Studies

The major studies can be compared as follows. The studies are presented in the same reference order used throughout this review.

Table 1: Comparative Analysis of Previous Studies

Ref.	Main focus	Dataset	Approach	Main reported result
[1]	Bengali digit recognition	NumtaDB	AlexNet, MobileNet, Inception V3, CapsuleNet	AlexNet: 99.01%
[2]	Digits and characters	CMATERdb, Ekush, BanglaLekha, NumtaDB	BengaliNet	Up to 99.13%
[3]	Multi-category character recognition	Ekush, CMATERdb	EkushNet	97.73% Ekush; 95.01% cross-validation
[4]	Bengali digit-recognition survey	Multiple datasets	ML/DL survey	Identified datasets, ambiguities and research directions
[5]	Numerals, basic, modifiers, compounds	CMATERdb, ISI, BanglaLekha, IUB datasets	RATNet	99.66% numerals; 97.70% CMATER compounds
[6]	Enlarged Bengali character space	Custom + Grapheme data	BNVGLENET	98.2% grouped task; ~97.5% large individual-class task
[7]	Unconstrained text recognition	2,338 lines, ~21K words	HOG + BLSTM-CTC	75.40% character accuracy
[8]	Handwritten document	CMATERdb1	Dataset/line	150 document

	dataset		segmentation	pages
[9]	Bengali word recognition	BanglaWriting	CNN + GRU/LSTM + CTC	CER 0.091; WER 0.273
[10]	Handwriting quality	322 English text lines	Structural features + SVM/k-NN	98.5% best test accuracy
[11]	Isolated basic characters	CMATERdb, ISI, BanglaLekha	BornoNet	Up to 98%
[12]	Bengali writer verification	100 writers	Modified FFT/GLCM/DCT	Handcrafted textural verification
[13]	CNN architecture comparison	Ekush	Eleven CNN architectures	ResNet152V2 among strongest models
[14]	Lexicon-free word recognition	IAM English dataset	YOLOv3 character spotting	Character-detection-based alternative
[15]	Cost-effective character recognition	Custom + CMATERdb	Compact CNN	95.67%; 0.8M parameters
[16]	Grapheme dataset/modeling	Bengali.AI	Multi-target grapheme representation	411K+ samples, 1,295 common graphemes
[17]	Bengali writer verification	100 writers	PDF-CNN + Siamese	2.36% best EER

There is a high level of performance gradient with the complexity of the tasks. The accuracy of isolated numeral recognition is often close to or equal to 100%, and for basic characters, it is also high. The recognition of compounds is typically weaker and the recognition of unconstrained lines is still significantly harder. The output is a variable length sequence and this is better reflected in Word-level recognition (CER, WER) than simple classification.

An even more significant distinction relates to computing needs. While some of the best-performing architectures make use of large deep networks, BengaliNet, BornoNet, RATNet, and the subsequent compact CNN provides evidence that competitive recognition can be accomplished with relatively small parameter numbers [2], [5], [11], [15].

2.8 Limitations of Existing Research

A problem with the early deep-learning research is that it has focused on isolated or limited tasks of recognition. While extremely high recognition accuracy of the digits could be achieved on standard CNN architectures [1], on pre-segmented images of numerals, they could not be achieved on images of handwritten digits. BengaliNet expanded the space of the class used but was still based on isolated symbols [2]. Another example of classification performed after separation of individual handwritten segments from words or documents is EkushNet [3].

The bigger body of Bengali handwriting research also exhibits substantial data-dependency. Different results for the same models are obtained from various sources CMATERdb, Ekush, BanglaLekha, and NumtaDB [4]. Compared to the results of other OCR systems, RATNet achieved significantly higher recognition accuracies for digits and basic characters than for compound characters and explicitly stated that its experiments were conducted on isolated characters, and needed another segmentation algorithm for doing full OCR [5]. Similar characters, labelling errors, writing problems and rotated samples were also found to cause errors in BNVGLENET [6].

Transitioning to handwriting will present a lot more challenges. In the case of Bengali language, BLSTM-CTC achieved only 75.40% character-level accuracy on unconstrained text and pointed out that post-processing can further enhance the accuracy [7]. It shows that even if the document lines are curved, skew, touching or mixed scripts, it is difficult to segment them into lines and words in the case of CMATERdb1. The end-to-end word recognition eliminates some of the segmentation dependency, but still gives a WER of 0.273 which still has a lot of room for improvement [9].

One more restriction is the lack of quality oriented handwriting studies in Bengali language. An in-depth study on English, not Bengali, in the selected literature studies was the strongest amongst dedicated hand writing quality studies [10]. In the case of Boronet, the focus was on recognition instead of on handwriting quality, and for Bengali, the focus was on writer identity and not on legibility or writing quality [11], [12]. The large CNN comparison also measured recognition without trying to determine if the character was well formed or structurally regular [13].

One approach is object-detection based word recognition, but the study mentioned was created for the English language and therefore is unable to directly address any new Bengali compound or diacritics [14]. The small-size Bengali CNN only focuses on the isolated classification [15]. However, grapheme level modeling is not a good choice for the representation of Bengali orthography, and was developed mainly for the recognition, not handwriting quality scoring [16]. In the same way, PDF-CNN-Siamese learning is able to learn writer-specific handwriting

patterns but cannot evaluate if the handwriting is good, bad, legible or structurally consistent [17].

In this way, the literature has progressed in the identification of the content of the text, and, to a certain degree, of the author itself, but it has not shown much interest in measuring the quality of the writing.

2.9 Research Gap

From the reviewed literature it is concluded that for the recognition of isolated handwritten characters, Bengali handwritten recognition has already reached a relatively mature stage. Deep CNNs can achieve very high accuracy for recognizing the Bengali digits [1] and efficient custom CNNs can recognize numerals, basic characters, modifiers and compounds on multiple benchmark sets [2], [3]. This is consistent with the overall trends of survey literature, which indicate a dramatic transition from handcrafted features to CNN-based approaches [4]. Several other works have also used residual-attention networks and enlarged CNN architectures to advance in the field of complex character recognition and to increase the number of characters to be recognized [5], [6].

But, the research decreases as it steps from isolated classification to natural handwriting. Recognizing unconstrained Bengali text is still significantly less accurate [7] and datasets containing document-level problems are shown with matra, modifiers, touching components and mixed scripts [8], which reveals segmentation issues that are yet to be resolved. The processing pipeline is improved by end-to-end word recognition, but it still has significantly higher errors than isolated word recognition systems [9].

However, above all, there is little research that is directly focused on the assessment of handwriting quality in Bengali. Characteristics like slant, spacing and size uniformity of characters can be quantified with good success, as demonstrated by existing handwriting-quality work, but the work selected in this study was for English handwriting [10]. The Bengali CNN systems are still concerned with the recognition problem [11] whereas the Bengali handwriting analysis systems are concerned with the writer verification problem and not handwriting quality [12]. Large studies on architecture comparison also focus on classification performance instead of interpretable measures [13].

Possible alternatives have been suggested by the methods proposed for other handwriting-recognition setups [14] and deployment of efficient Bengali CNN has shown that it is possible to deploy the system using compact CNN architecture [15]. Therefore, grapheme-based representation has more linguistically suitable basis for modelling Bengali roots and diacritics [16] and hybridization between handcrafted and learned approaches shows the

possibility of combining structural handwriting with learned representations successfully [17]. But these developments are not yet linked with a specific Bengali handwriting-quality system.

So, one of the key research gaps is that there is no automated handwriting quality assessment system available in Bengali language that quantitatively assesses script-relevant structural characteristics and utilizes the modern machine-learning or deep-learning representations. Such a framework should include the recognition accuracy and other aspects, such as character size uniformity, spacing uniformity, slant, alignment, matra uniformity, proportions of the various writing areas, placement of modifiers, and formation of characters or graphemes.

A good system would, then, be to extend the current research question to: What Bengali character, word or writer is this handwriting?

To: Is this Bengali hand-writing good, consistent and structurally correct?

The difference is the main gap that was found after conducting a literature review, and gives a better understanding of where to go next for more research.

3. Theoretical Foundations

In this chapter the theory is presented which is the basis of the proposed Bengali handwriting quality assessment framework. It brings in the understanding of digital image representation, normalization of digital images, resizing and Convolutional Neural networks that are essential for processing handwriting images. In addition, the chapter covers convolution, activation functions, pooling, batch normalization, dropout, global average pooling and Softmax classification. The focus is on multi-resolution feature learning, in which the fine-grained, mid-level, and global features of handwriting are extracted at different spatial resolutions. Feature fusion, transfer learning, optimization with Adam, categorical cross-entropy loss and standard evaluation metrics are also discussed. These ideas are used in the development of the architecture and experimental approach detailed in the following chapters.

3.1 Digital Image Processing

This section presents the theoretical foundations underlying the proposed Bengali handwriting quality assessment system. The theoretical framework encompasses digital image processing fundamentals, convolutional neural network architectures, multi-resolution feature learning principles, feature fusion strategies, transfer learning concepts, optimization techniques, and evaluation metrics. Understanding these theoretical concepts is essential for comprehending the design decisions and implementation details presented in subsequent chapters.

3.1.1 Digital Image Representation

A digital image is a two-dimensional array of intensities of the pixels in the image. In a grayscale image, every pixel holds a single value that represents the intensity of the image at that point, a value between 0 (black) and 255 (white). In the case of color images, each picture element contains a set of three numbers: the Red number, the Green number and the Blue number**.**

$$I(x,y) = \begin{cases} f(x,y) & \text{for grayscale images} \\ [f_R(x,y), f_G(x,y), f_B(x,y)] & \text{for RGB images} \end{cases} \quad \textbf{(I)}$$

Where :
I(x, y) = Pixel intensity at spatial coordinates (x, y).
f(x, y) = Intensity value for grayscale images [0, 255].
f_R, f_G, f_B = Intensity values for red, green, and blue channels [0, 255].
(x, y) = Spatial coordinates in the image plane.
The image for the handwriting quality assessment task is encoded as a tensor of height H, width W and channels RGB of shape H × W × 3.

3.1.2 Image Normalization and Resizing

A technique used in image processing which involves transforming pixel intensities from one range to another, usually [0, 1] or [-1, 1] is called image normalization. This normalization allows stable gradient propagation in training and uniform distribution of inputs to all samples.

$$x_{normalized} = \frac{x_{original} - x_{min}}{x_{min} - x_{min}} = \frac{x_{original}}{255.0} \quad \textbf{(II)}$$

Where:

$x_{original}$ = Original pixel intensity value (0-255).

$x_{normalized}$ = Normalized pixel intensity value (0-1).

x_{min} = Minimum possible pixel value (0).

x_{max} = Maximum possible pixel value (255).

The image resizing operation resizes all images to a fixed size to enable the input tensor size to be the same for all images. The resize operation will use bilinear or bicubic interpolation for image quality:

$$I_{resized}(x', y') = \sum_{i=0}^{n} \sum_{j=0}^{n} w_i w_j I(x + i, y + j) \quad \textbf{(III)}$$

Where:

$I_{resized}$ = Resized output image.

(x', y') = Coordinates in the resized image.

(x, y) = Corresponding coordinates in the original image.

w_i, w_j = Interpolation weights.

n = Interpolation kernel size.

3.2 Convolutional Neural Networks

A type of deep learning architecture that was created to process grid-like data like images is called Convolutional Neural Networks (CNNs). CNNs apply local connectivity, parameter sharing and hierarchical feature learning to obtain state-of-the-art performance in visual recognition tasks.

3.2.1 Convolution and Feature Maps

The Convolution operation is the building block of a CNN which is used to extract the spatial features from the input image. Each filter (kernel) is a learnable parameter applied to the input, going over the spatial dimensions to create a feature map:

$$(F * K)(i,j) = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} F(i+m, j+n) \cdot K(m,n) + b \quad \textbf{(IV)}$$

Where:

F = Input feature map.

K = Convolutional kernel (filter).

k = Kernel size.

(i, j) = Spatial position in the output feature map.

b = Bias term.

The output size of a convolution operation is determined by the input size, kernel size, stride, and padding:

$$O = \frac{W - K + 2P}{S} + 1 \quad \textbf{(V)}$$

Where:
O = Output spatial dimension.
W = Input spatial dimension.
K = Kernel size.
P = Padding.
S = Stride.

3.2.2 Activation and Pooling

Activation Functions: These add non-linearity to the network, allowing the learning of complex patterns. The most commonly used activation function is the Rectified Linear Unit (ReLU).

$$f(x) = \max(0, x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \quad \textbf{(VI)}$$

ReLU has a number of beneficial properties such as sparsity, computational efficiency and the elimination of the vanishing gradient problem.

Pooling operations: Pooling layers decrease the spatial size of feature maps and introduce translation invariance and decrease computational complexity. The most frequently used pooling operation would be max pooling.

$$P(i,j) = \max_{m=0}^{k-1} \max_{n=0}^{k-1} F(i \cdot S + m, j \cdot S + n) \quad \textbf{(VII)}$$

Where:
P = Pooled output.
F = Input feature map.
k = Pooling window size.
S = Stride.
(i, j) = Spatial position in the pooled output.

3.2.3 Batch Normalization and Dropout

Batch Normalization: Batch normalization normalizes the activations of a layer over the batch dimension to help train faster and more stable.

$$\hat{x} = \frac{x - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, \quad y = \gamma \hat{x} + \beta \quad \textbf{(VIII)}$$

Where:
x = Input to the batch normalization layer.
P_B = Batch mean.
g_{B2} = Batch variance.
E = Small constant for numerical stability.
y = Scale parameter (learnable).
B = Shift parameter (learnable).
y = Normalized output.

Dropout: A regularization technique that randomly turns off a subset of neurons during training, to avoid co-adaptation and to decrease overfitting.

$$y = \begin{cases} 0 & \text{with probability } p \\ \frac{x}{1-p} & \text{with probability } 1-p \end{cases} \quad \textbf{(IX)}$$

Where:
x = Input activation.
y = Output after dropout.
p = Dropout rate (probability of deactivation).

3.2.4 Global Average Pooling and Softmax

Global Average Pooling (GAP): GAP reduces the spatial dimensions of feature maps to 1×1 by computing the average across each feature map.

$$v_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_c(i,j) \quad \textbf{(X)}$$

Softmax Activation: Softmax converts raw logits into probability distributions over classes.
Softmax Function:

$$\hat{y}_c = \frac{e^{z_c}}{\sum_{j=1}^{C} e^{z_j}} \quad \textbf{(XI)}$$

Where:
$\hat{y}_c$ = Predicted probability for class c.
z_c = Logit score for class c.
C = Total number of classes (5).

3.3 Multi-Resolution Feature Learning

The main theoretical concept behind the proposed architecture is the multi-resolution feature learning. The method is based on the idea that various visual aspects of handwriting are more effectively captured at different spatial resolutions, similar to how human evaluators determine the quality of handwriting at various resolutions.

3.3.1 Fine-Grained Features

Fine-grained features capture local, high-frequency information at the pixel and sub-pixel level. In handwriting quality assessment, these features include:
Stroke Thickness: Consistency and uniformity of line widths.
Edge Sharpness: Clarity and definition of stroke boundaries.
Texture Patterns: Surface characteristics of the writing.
Local Irregularities: Tremors, breaks, and inconsistencies in small regions.

Fine-Grained Feature Extraction:

$$F_{fine} = f_{conv}^{(l)}(x), \quad l \in \{1,2,3\} \quad \textbf{(XII)}$$

3.3.2 Mid-Level Structural Features

Mid-level features extract patterns at an intermediate level of granularity that can be observed from the composition of individual strokes. These features include:
Curvature Patterns: Trajectories of curved strokes.
Stroke Junctions: Connection points between different strokes.
Local Relationships: Spatial arrangements of stroke elements.
Compound Character Structures: Merged character formations.

Mid-Level Feature Extraction:

$$F_{mid} = f_{conv}^{(m)}(x), \quad m \in \{4, 5, 6\} \qquad \textbf{(XIII)}$$

Where:
F_{mid} = Extracted mid-level features.
$f_{conv}^{(m)}$ = Convolutional layers at intermediate resolution.
x = Input image.

3.3.3 Global Shape Features

Global features capture low frequency information, overall character structure and layout. These features include:
Character Proportions: Height-to-width ratios and element sizing.
Spatial Layout: Distribution of elements within the character.
Overall Alignment: Vertical and horizontal consistency.
Global Symmetry: Balance and symmetry of the character.
Global Feature Extraction:

$$F_{coarse} = f_{conv}^{(h)}(x), \quad h \in \{7, 8, 9\} \qquad \textbf{(XIV)}$$

Where:
F_{coarse} = Extracted global features.
$f_{conv}^{(h)}$ = Convolutional.

3.4 Feature Fusion

Feature fusion is the merging of complementary features from various sources to form a more complete representation. The proposed architecture is based on channel-wise concatenation of features obtained from the three parallel streams.

Feature Fusion via Concatenation:

$$F_{fused} = [F_{fine} \oplus F_{mid} \oplus F_{coarse}] \quad \textbf{(XV)}$$

Where:
F_{fused} = Fused feature representation.
F_{fine} = Fine-grained features.
F_{mid} = Mid-level features.
F_{coarse} = Coarse global features.
$\oplus$ = Channel-wise concatenation operation.

Fused Feature Dimensions:

$$C_{fused} = C_{fine} + C_{mid} + C_{coarse} \quad \textbf{(XVI)}$$

Where:
C_{fused} = Total channels in fused representation.
C_{fine} = Channels from fine-grained stream.
C_{mid} = Channels from mid-level stream.
C_{coarse} = Channels from coarse stream.

3.5 Transfer Learning

Transfer learning is a machine learning method in which learning from a specific task can be used to boost the performance of a similar task. Transfer learning in computer vision is usually a process that begins by training a model on a massive image dataset (like ImageNet) and then fine-tuning it on the specific dataset of interest.

Transfer Learning Fine-Tuning:

$$\theta^* = \arg\min_{\theta} \mathcal{L}(f_{\theta_{pre-trained}}(x), y) \quad \textbf{(XVII)}$$

Where:
θ^* = Optimized parameters after fine-tuning.
$\theta_{pre\text{-}trained}$ = Parameters from pre-trained model.
$\mathcal{L}$ = Loss function.
f = Model function.
x = Input image.
y = Ground truth label.

3.6 Optimization and Loss Function

Optimization techniques are used to minimize the loss function and determine the optimal parameters of the model. Optimizer and loss function greatly affect the process of training and the quality of the trained model.

3.6.1 Adam Optimizer

Adam (Adaptive Moment Estimation): An adaptive learning rate optimization algorithm that is a combination of RMSprop and momentum.

Adam Optimizer Updates:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\nabla L_t \quad \textbf{(XVIII)}$$

Second Moment Estimate:

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)(\nabla L_t)^2 \quad \textbf{(XIX)}$$

Bias-Corrected Estimates:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad \textbf{(XX)}$$

Parameter Update:

$$\theta_{t+1} = \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad \textbf{(XXI)}$$

Where:
θ_t = Parameters at time step t.
∇L_t = Gradient of loss at time step t.
m_t, v_t = First and second moment estimates.
β_1, β_2 = Exponential decay rates (0.9, 0.999).
a = Learning rate ($1x10^{-3}$).
$\in$ = Small constant (1×10^{-7}).

3.6.2 Categorical Cross-Entropy

Categorical cross-entropy is the measure of difference between the actual probability distribution and predicted probability distribution

Categorical Cross-Entropy Loss:

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c}) \quad \textbf{(XXII)}$$

Where:
N = Total number of training samples.
C = Total number of classes (5).
$y_{(i,c)}$ = Ground truth label (1 if sample i belongs to class c, else 0).
$\hat{y}_{(i,c)}$ = Predicted probability that sample i belongs to class c.

3.7 Evaluation Metrics

Evaluation metrics give a numeric score of the models' performance, allowing for an objective comparison of the different architectures and approaches [41-52].

3.7.1 Accuracy, Precision and Recall

Accuracy measures the overall proportion of correctly classified samples:

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} 1[\hat{y}_i = y_i] \quad \textbf{(XXIII)}$$

Precision quantifies the reliability of positive predictions for each class:

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c} \quad \textbf{(XXIV)}$$

Where:
TP_c = True positives for class c.
FP_c = False positives for class c.
FN_c = False negatives for class c.

Recall measures the model's ability to find all samples belonging to a class:

$$\text{Recall}_c = \frac{TP_c}{TP_c + FN_c} \quad \textbf{(XXV)}$$

Where:
N = Total number of test samples.
$\hat{y}_i$ = Predicted class for sample i.
y_i = True class for sample i.
1[-] = Indicator function (1 if true, O otherwise).

3.7.2 *F_1-Score*

F_1-Score is a balanced measure of model performance that combines both precision and recall by using their harmonic mean:

$$\text{F1}_c = 2 \times \frac{\text{Precision}_c \times \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \quad \textbf{(XXVI)}$$

The harmonic mean is used instead of the arithmetic mean since it will penalise extreme imbalance between precision and recall, thus only good models will be considered good.

3.7.3 Confusion Matrix

The confusion matrix gives a detailed performance summary for each class at a time. $A_{(i,j)}$ is the number of samples in class i that were classified into class j:

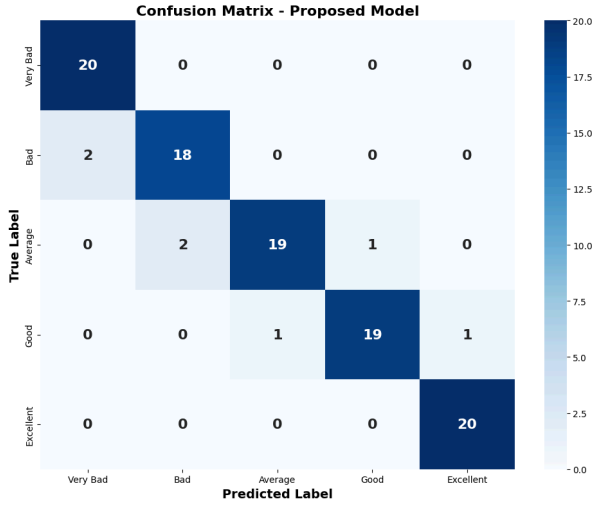


Figure 1: Confusion Matrix of the proposed model.

Confusion Matrix:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1C} \\ a_{21} & a_{22} & \cdots & a_{2C} \\ \vdots & \vdots & \ddots & \vdots \\ a_{C1} & a_{C2} & \cdots & a_{CC} \end{bmatrix} \quad \textbf{(XXVII)}$$

Where:
$a_{(i,j)}$ = Number of samples from true class i predicted as class j.
C = Total number of classes (5).

The diagonal elements $a_{(i,j)}$ represent correct classifications, while off-diagonal elements represent misclassifications.

4. Dataset Preparation and Preprocessing

This chapter presents the preparation of the dataset which is used for training and testing the Bengali handwriting quality assessment system. In this research work, 1000 Handwritten Bengali character images are used, which are equally distributed in five different quality categories: Very Bad, Bad, Average, Good and Excellent. The collected samples were representative of natural handwriting and evaluated for stroke quality, structural correctness, legibility, and overall consistency. In this dataset the data was split in to training, validation and testing data sets in the ratio of 80:10:10. All standard image processing was performed prior to model training such as image resize, pixel normalization and grayscale to RGB. Rotation, zoom and flipping were also applied to data augmentation to boost variation and model generalisation.

4.1 Dataset Overview

In this work, a dataset of 1000 handwritten images of Bengali characters, which are carefully selected and annotated for five different quality conditions (Very Bad, Bad, Average, Good, Excellent) is used. There are 200 samples in each category with an equal number of samples in each quality level. The data set was constructed to reflect the natural variability that exists in actual handwriting, such as variations in stroke quality, character formation, legibility, and overall consistency.

The dataset is used for training and testing the proposed multi-resolution CNN architecture for automated handwriting quality assessment. This is in contrast to the current dataset which has mainly concentrated on character recognition (what was written) and this dataset concentrates specifically on handwriting quality evaluation (how well it was written) which is an under-explored problem. This exclusive focus allows it to be valuable for educational assessment, forensic application and archival digitization.

4.2 Data Collection Process

A variety of samples to represent real world situations were collected to capture a wide range of handwriting samples. The sample characters were taken in handwritten form from various sources to cover extensive writing variations of Bengali characters. Figure 2 presents representative samples from each quality class.

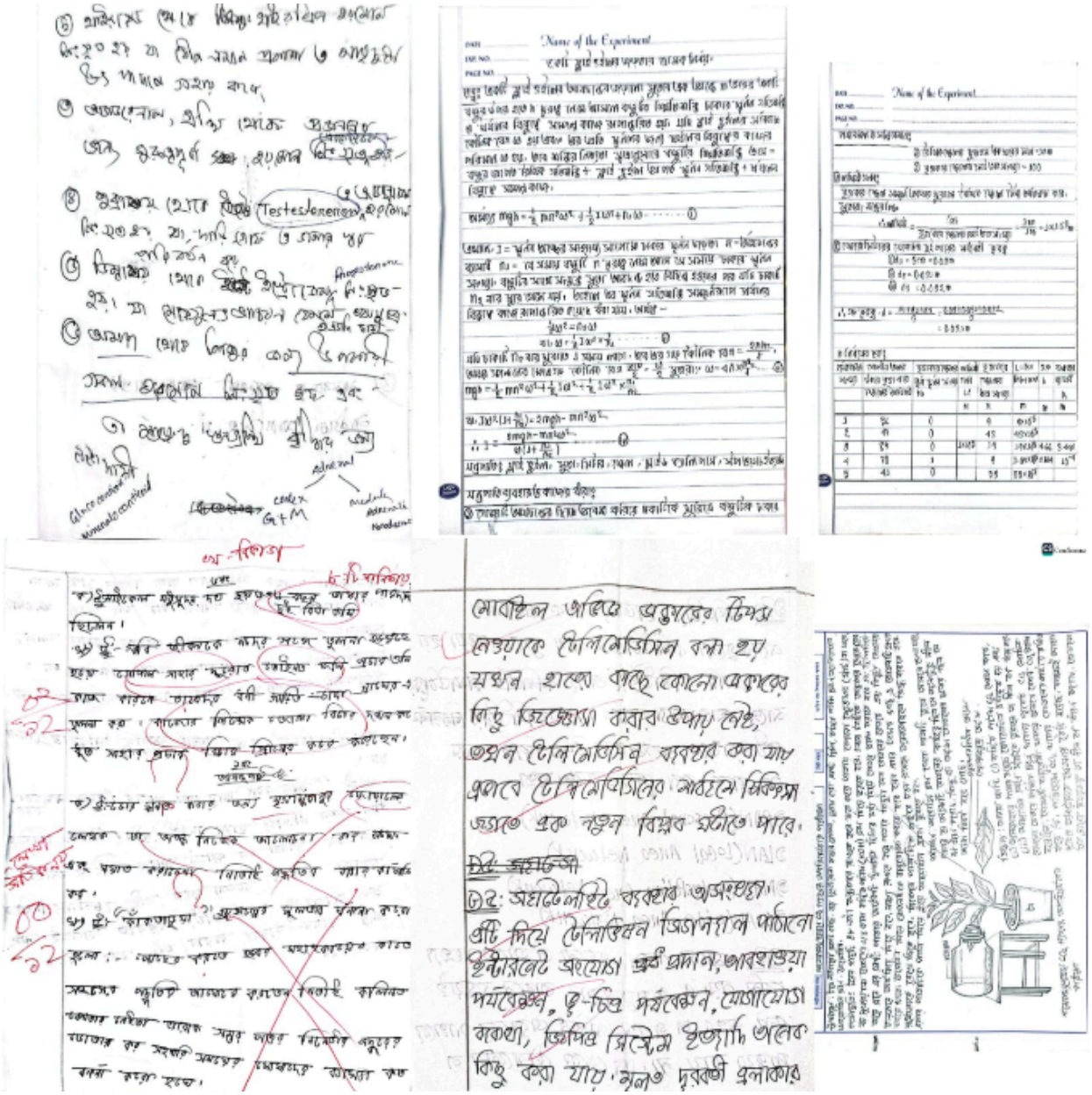


Figure 2: Data collection process.

Figure 2 shows typical examples of each of the five quality categories. The images are clearly marked to show the development from the Very Bad (poor stroke quality, inconsistent shapes, irregular proportions) to Excellent (smooth strokes, consistent shapes, proper proportions). This image illustrates the problem of ambiguity with handwriting quality assessment: minute variation of the quality class between adjacent quality classes leads to difficulty in classification.

4.2.1 Collection Sources

The data set was created from the following sources:
1.Educational Institutions: Students in various grade levels were selected at educational institutions, covering a wide range of handwriting skills, from beginners to advanced writers.
2.Community Contributors: Samples were collected from members of the community who had a variety of writing experiences and include writers as well as students and people who have different writing habits.
3.Public Handwriting Databases: A few samples were retrieved from publicly available handwriting databases which were diverse in writing style and character variation.
4.Controlled Collection: A sub-sample of samples was taken under controlled conditions to produce samples to ensure balanced representation of each quality level.

4.2.2 Collection Specifications

The photographs gathered were all of the following dimensions:
1.Image Format:JPEG/PNG.
The minimum image resolution is variable, but it must be at least 224×224.
2.Writing Instrument: Pen / Pencil (Various Types).
3.Writing Surface: Paper (various qualities).
4.Character Type: Bengali individual characters and simple words.
5.Background: No background (be sure to avoid background interference).

4.3 Handwriting Quality Classes

Handwriting samples were classified into five handwriting quality classes representing different levels of handwriting skills. The classification system was designed to encompass the full range of the quality of handwriting, from very poor handwriting to very good handwriting.

4.3.1 Very Bad and Bad

Very Bad (Class 0): These samples show a very poor handwriting. Characteristics include:
Quality of stroke: Very irregular stroke thickness, interrupted or incomplete strokes and shivering strokes.
- Shape and Structure: Excessively misproportioned characters, unusual shapes and forms.
- Legibility: Very hard to read/readability is 100% impossible.
- Consistency: Very inconsistent in one or more characters.
- ***Consistency:***Highly inconsistent within a single character and across multiple characters.

Bad (Class 1): Poor handwriting, definitely not up to standard. Characteristics include:
- Stroke Quality: Thickness of strokes is not uniform and some strokes may be broken and wavy.
- Shape and Structure: Some character shape and some distortion, disproportionate body elements, some recognition.
- Readability: Not readable but can be recognized with effort.
- Consistency: Notable lack of consistency within and between characters.

4.3.2 Average

Average (Class 2): Acceptable but not outstanding quality of handwriting. This category is the middle ground and can often cause the most classification problems because it is the most ambiguous. Characteristics include:
- Weakness: Generally good quality of stroke with occasional irregularities.
- Shape and Structure: Mostly correct character shapes, some proportional errors.
- Legibility: Generally clear with a little effort, but some characters may be interpreted.
- Consistency: Some consistency with some variation.

4.3.3 Good and Excellent

Good (Class 3): These are samples of well-written characters that do not have many flaws. Characteristics include:
- Stroke Quality: Regular stroke thickness, clear lines and sharp edges.
- Proportions: Accurate size and proportions of characters.
- Readability: Readability without effort.
This is a great way to test consistency.This is a good consistency test: Consistency: High consistency with minimal variation.

Excellent (Class 4): These samples are the best quality of handwriting and they demonstrate good standards of writing. Characteristics include:
- Stroke Quality: Perfectly consistent stroke thickness, exceptionally smooth and clean lines.
- Shape and Structure: Well-formed characters, well-proportioned, with perfect structural elements.
- Legibility: Readable without any difficulty or doubt.
- Consistency: Very good consistency throughout.

4.4 Annotation and Quality Assessment Criteria

Human experts having experience in handwriting assessment, educational evaluation and calligraphy have performed the annotation process. A uniform quality assessment system was established to guarantee that each sample is annotated in a reliable and impartial manner.

Table 2: Annotation and Quality Assessment

Aspect	Very Bad (0)	Bad (1)	Average (2)	Good (3)	Excellent (4)
Stroke Quality	Broken, inconsistent, tremulous	Irregular, shaky	Generally consistent	Clean, consistent	Flawless, smooth
Shape Correctness	Severely distorted	Moderate distortion	Minor errors	Correct shapes	Perfect formation
Structural Integrity	Unrecognizable forms	Partial recognition	Mostly recognizable	Clear structure	Ideal structure
Legibility	Illegible	Difficult to read	Readable with effort	Easy to read	Effortlessly clear
Consistency	Highly inconsistent	Noticeable inconsistency	Moderate consistency	High consistency	Exceptional consistency

Table 2 shows the full form of the quality assessment criteria that were applied during annotation. The average rating for each aspect was used to calculate the overall quality class and each aspect was rated on the scale of 0-4. The criteria were developed to reflect the multi-dimensional nature of handwriting quality and provide a comprehensive assessment.

4.4.1 Stroke Quality

Stroke quality assessment focused on evaluating the physical characteristics of the written strokes. Key indicators included:

- *Stroke Thickness Consistency:*Whether the stroke width remains uniform throughout the character

- *Stroke Smoothness:*The absence of tremors, breaks, or irregular edges

- *Stroke Direction:*Proper directionality of strokes according to character formation rules

- *Line Quality:*The clarity and definition of individual strokes

4.4.2 - Shape and Structural Correctness

Shape and structural correctness assessment: Geometric and structural properties of characters were assessed.

These items represent a good balance between height and width proportions and size of the elements.

- Curvature Quality: Smooth and suitable curvature in curved elements.

- Junction Quality: Good connections between the strokes.

- Matra Line Consistency: For Bengali characters the line consistency and position of the matra line.

4.4.3 Overall legibility and consistency

Legibility and consistency assessment investigated the overall legibility and uniformity:
- Character Recognition: Easy to recognize intended character.
- Efficient Use of Space: Overall use of space efficiency.
- Spacing between words or between words and symbols: Correct Spacing.
- Uniformity of the characters from the same author (Global Consistency).

4.5 Dataset Distribution and Split

In order to avoid any biases in the development and evaluation of the model, the data set was systematically divided into three mutually exclusive subsets: train, validation and test sets. The 80:10:10 split ratio was chosen because it was deemed to balance the amount of data needed for training with the amount needed for testing and validation.

4.5.1 Training Set

The training set is 80% of the whole data set (800 images). This subset is used as the main learning material for the models.
Key characteristics include:
- Size: 800 images (160 per class).
Using neural networks to model weight optimization and feature learning.Modeling weight optimization and feature learning using neural networks.
- Augmentation: Augmented during training to help generalize.
- Balance: Evenly distributed by all 5 quality classes.

4.5.2 Validation and Test Sets

Documenting a validation set and test set.
Each set (validation and test) consists of 10% of the whole set (100 images).
- Validation Set (10%):100 images (20 per class) - For hyperparameter tuning and early stopping.
- Test Set (The last 10%):100 images (20 per class) - For final model evaluation and performance reports.

Table 3: Validation and Test Sets

Quality Class	Total	Train (80%)	Validation (10%)	Test (10%)
Very Bad	200	160	20	20
Bad	200	160	20	20
Average	200	160	20	20
Good	200	160	20	20
Excellent	200	160	20	20
Total	**1000**	**800**	**100**	**100**

It shows the complete data set distribution is given for all five quality classes in Table 3. The split ratio of 80:10:10 guarantees that the training and validation sets are sufficiently large, while also providing a good balance in the test set. This equilibrium distribution allows for a non-biased model development and evaluation.

4.6 Image Preprocessing

All images were preprocessed using a consistent pipeline to ensure that they are of a uniform quality and suitable for feeding into the convolutional neural network architectures used for training. The pre-processing steps are aimed at preserving the meaningfulness of the features relevant to the quality, while normalizing image properties.

4.6.1 Image Resizing

To ensure that all tensors from the model have the same shape as input, all images were resized to a fixed size of 224 x 224 pixels. This dimension was chosen because it is a standard size input that can be handled by all baseline architectures, but also has enough resolution for the multi-resolution streams.

$$I_{resized} = \text{Resize}(I_{original}, (224, 224)) \quad \textbf{(XXVIII)}$$

Where:
$I_{original}$ = Original input image of arbitrary dimensions.
$I_{resized}$ = Resized image of fixed dimensions 224×224 pixels.
Resize() = Bicubic interpolation resizing function.

4.6.2 Pixel Normalization

The pixel intensities were rescaled from [0, 255] to [0, 1] to ensure that the gradient will propagate during training. This normalisation provides uniform distribution of inputs and faster convergence.

Pixel Normalization:

$$x_{normalized} = \frac{x_{original}}{255.0} \quad \textbf{(XXIX)}$$

Where:

$x_{original}$ = Original pixel intensity value (0-255).

$x_{normalized}$ = Normalized pixel intensity value (0-1).

4.6.3 Grayscale-to-RGB Conversion

Images were also converted from grayscale to three channels (RGB) to be compatible with pre-trained model architectures which need three channels. This conversion process is identical to the single channel on all three RGB channels:

$$I_{RGB} = [I_{gray}, I_{gray}, I_{gray}] \quad \textbf{(XXX)}$$

Where:

I_{gray} = Single-channel grayscale image.

I_{RGB} = Three-channel RGB image.

4.7 Data Augmentation

To enhance the generalization of the model and mitigate overfitting, data augmentation techniques were employed during the training process. The augmentation pipeline was designed carefully to augment the training data with realistic variations while preserving the quality-relevant features.

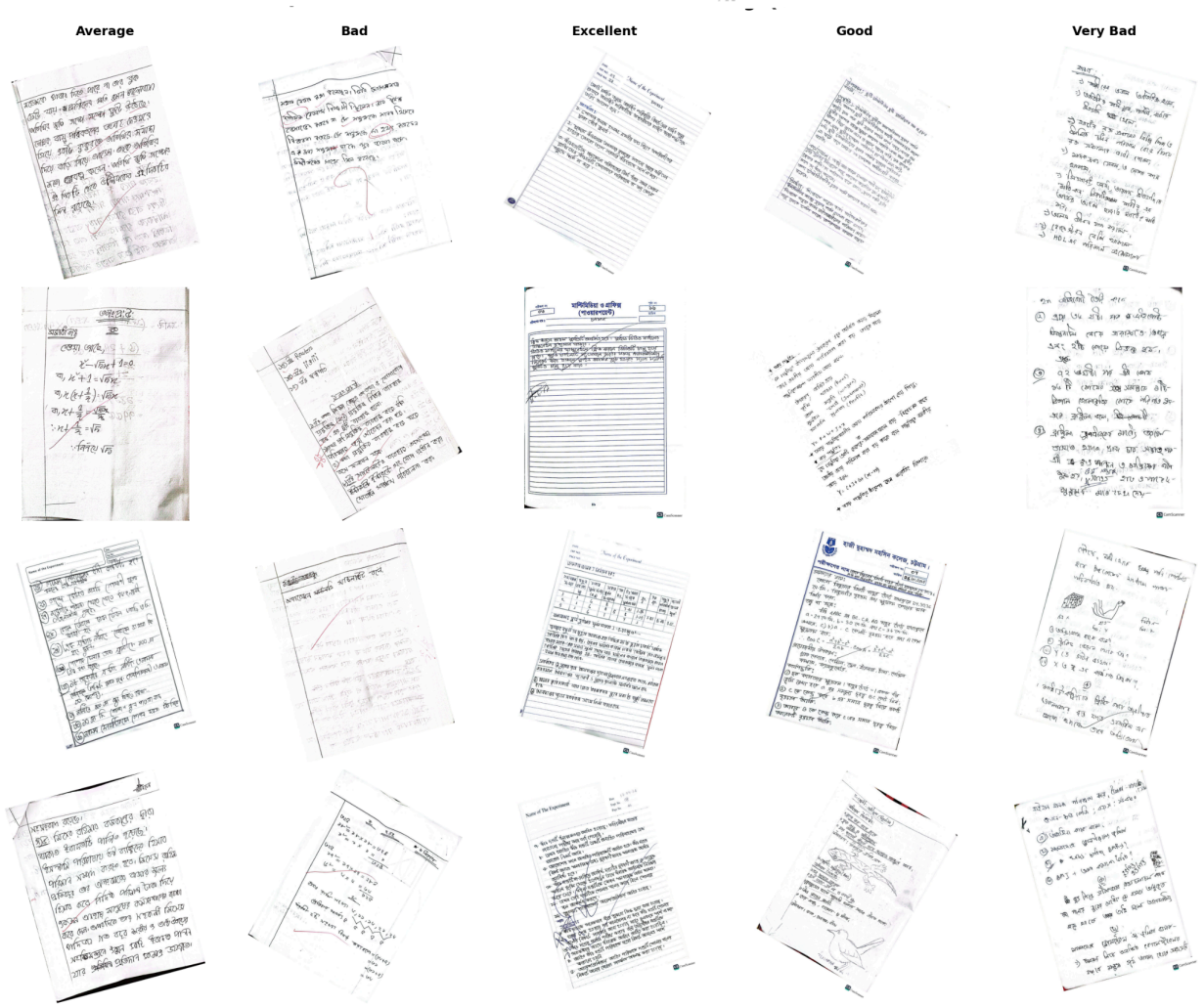


Figure 3: Data augmentation.

The different augmentation methods used on the training data are shown in Figure 3. The images in each row are the same image, but with different augmentations: rotation, horizontal flip, and zoom. These enhancements duplicate the variations of handwriting presentation in the real world, but maintain the essential qualities.

4.7.1 Rotation

Random writing angle variation was added by random rotation to simulate variations in writing angle that occur naturally in handwriting. The rotation operation is given by:

$$I_{rotated} = \text{Rotate}(I_{original}, \theta), \quad \theta \sim \mathcal{U}(-20^\circ, 20^\circ) \tag{XXXI}$$

Where:
$I_{original}$ = Original input image.
$I_{rotated}$ = Rotated image.
θ= Rotation angle uniformly sampled from [-20°, 20°].
U = Uniform distribution.
To avoid the possible distortion that could affect quality assessment features, rotation angles were constrained between ±20°.

4.7.2 Zoom and Flipping

Zoom Augmentation: To simulate different capture distances and different image scales a zoom augmentation was applied:

$$I_{zoomed} = \text{Zoom}(I_{original}, s), \quad s \sim \mathcal{U}(0.8, 1.2) \tag{XXXII}$$

Where:
$I_{original}$ = Original input image.
I_{zoomed} = Zoomed image.
s = Zoom factor uniformly sampled from [0.8, 1.2].
U= Uniform distribution.
Zoom factors were selected to simulate both close-up and distant capture scenarios while maintaining character visibility.

Horizontal Flipping: Applied with the probability of 0.5 to improve the symmetry learning and helps model learn features invariant to horizontal orientation, improving generalization:

$$I_{flipped} = \begin{cases} \text{FlipHorizontal}(I_{original}) & \text{with probability } p = 0.5 \\ I_{original} & \text{with probability } p = 0.5 \end{cases} \tag{XXXIII}$$

Where:
$I_{original}$ = Original input image.
$I_{flipped}$ = Horizontally flipped image.
p = Probability of applying flip (0.5).

5. Proposed Methodology

In this chapter the proposed methodology and architectural design of the proposed system for automated Bengali handwriting quality assessment is presented. The overall structure of the framework is a systematic pipeline that involves data preparation, preprocessing, splitting the dataset, training the model, evaluating the performance and comparison of models. The proposed CNN is at the heart three parallel feature extraction streams that are able to extract handwriting information at various spatial resolutions. The fine-grained stream deals with details of stroke and edge, the mid-grain stream deals with the pattern of curves and junctions, and the coarse stream may deal with the character shape and spatial arrangement. Features from these streams are fused together and finally classified into five handwriting quality levels by feature fusion. The proposed architecture is also compared to other architectures such as EfficientNetB0, ResNet50, InceptionV3, and Xception under common experimental conditions.

5.1 Overall Research Framework

The development of the proposed research framework is presented in a systematic pipeline for automatic Bengali handwriting quality assessment as shown in methodology flowchart Figure 4 . The framework starts with the collection of Bengali samples with handwritten characters from various sources to represent the natural variation in writing style and levels. The Image Pre-Processing stage involves applying a series of standardized transformations such as resizing, normalization, augmentation, etc., to the collected data to prepare it for the model training process. The Data Splitting phase is used for dividing the preprocessed data set into the Training set (80%), Validation set (10%), and Test set (10%) without any bias. At the Model Training stage, five different CNN architectures are implemented, including the proposed multi-resolution model and four baseline models: ***EfficientNetB0***, ***ResNet50***, ***InceptionV3***, ***and Xception***, all of which are trained under the same conditions. In the Performance Evaluation phase, each model is evaluated with several complementary metrics and in Result Analysis and Comparison, the best architecture is chosen. Lastly, the framework ends with the Conclusion phase that summarizes the results and suggests the best methods for handwriting quality assessment.

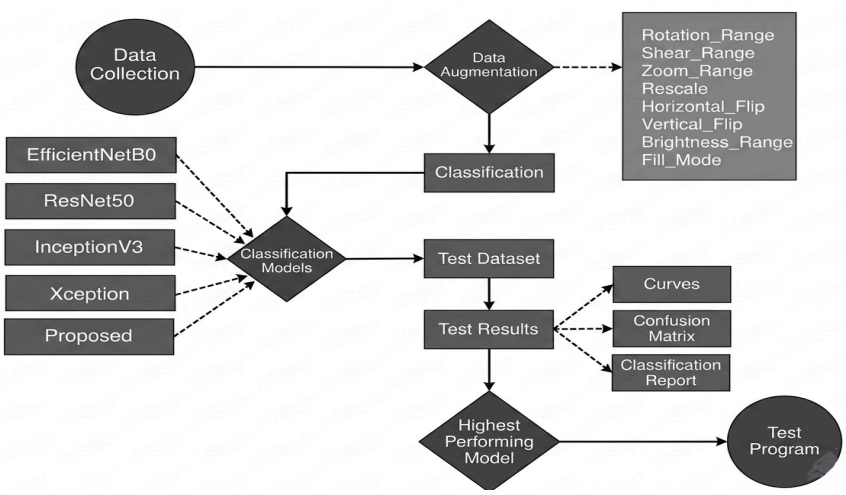


Figure 4: Research Framework.

Figure 4 illustrates the complete workflow of the proposed methodology. It begins with Image Collection, followed by Image Pre-Processing (Noisy Image Removal and Data Augmentation), Data Splitting (Train/Test/Validation), Model Training (Proposed Model and Baseline Models), Performance Evaluation, Result Analysis and Comparison, and finally the Conclusion stage. This structured pipeline ensures systematic development and evaluation of the handwriting quality assessment system.

5.2 Dataset Description and Preprocessing

The data set used in this research is the Bengali handwritten character images which are taken from various sources and manually annotated for quality assessment. The data set contains 5 different classes which represent different levels of **handwriting quality, Very Bad, Bad, Average, Good, and Excellent**. A representative sample of images for each quality class is shown in Figure 5.

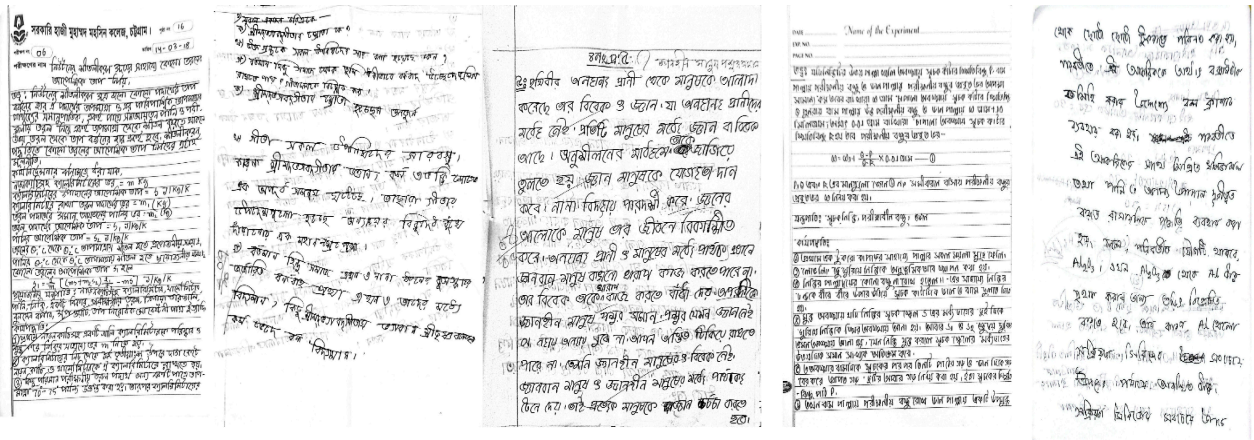


Figure 5: Dataset description and preprocessing.

Figure 5 displays representative samples from each of the five quality categories. The images clearly illustrate the progression from Very Bad (poor stroke quality, inconsistent shapes, irregular proportions) through to Excellent (smooth strokes, consistent shapes, proper proportions). This visual representation demonstrates the inherent

challenges in handwriting quality assessment, where subtle differences between adjacent quality classes create classification ambiguity.

5.2.1 Dataset Splitting Strategy

The data set was then systematically split into three subsets to avoid the bias in model development and evaluation. The detailed distribution of samples in all quality classes is given in Table 4.

Table 4: Dataset Splitting Strategy

Quality Class	Total	Train (80%)	Validation (10%)	Test (10%)
Very Bad	200	160	20	20
Bad	200	160	20	20
Average	200	160	20	20
Good	200	160	20	20
Excellent	200	160	20	20
Total	**1000**	**800**	**100**	**100**

Table 4 shows the complete dataset distribution across all five quality classes. The 80:10:10 split ratio ensures that each class has sufficient samples for training while maintaining balanced representation in validation and test sets. This balanced distribution prevents bias during model development and evaluation.

5.2.2 Preprocessing Pipeline

All images were preprocessed in a standardized pipeline prior to training to provide consistent input for training. The input images were scaled to have a fixed dimension of 224 × 224 pixels to make the tensors' shapes identical for all images. The intensity values of the pixels were then normalized between 0 and 1 with the following transformation:

$$x_{normalized} = \frac{x_{original}}{255.0} \quad \textbf{(XXXIV)}$$

Where:
$x_{original}$ = Original pixel intensity value (0-255).
$x_{normalized}$ = Normalized pixel intensity value (0-1).

5.2.3 Data Augmentation

DTo ensure generalisation across the data set and minimise overfitting, data augmentation techniques were used during training.These include:
1.**Random Rotation**: ±20 degrees to simulate slight writing angle variations.

2.**Horizontal Flipping**:With a probability of 0.5 to enhance symmetry learning.
3.**Zoom Augmentation**: Scaling factors uniformly sampled from [0.8, 1.2].
Images that have been originally captured as grayscale were converted to three-channel RGB, to be compatible with the pretrained model architectures that require three-channel inputs.

5.3 Proposed Multi-Resolution CNN Architecture

The suggested Multi-Resolution Feature Integration framework is particularly developed for Bengali Handwritten Quality Evaluation. This architecture is designed for handwriting quality assessment and is different from generic architectures trained with natural images, because it assumes that assessment of handwriting needs multi-scale analysis of both local and global handwriting properties, such as stroke structure and handwriting shape. When human users assess penmanship, they are able to combine information from these scales, such as consistency of strokes and overall alignment and proportion of characters.
The architecture consists of 3 parallel convolutional streams with different spatial resolutions as shown in Figure 6. Dedicated convolutional blocks are applied to each stream of the input image, and the architecture is different for each block in order to capture different scale features. The Fine-Grained Stream is at a very fine level of resolution, encoding stroke-level features, the Mid-Level Stream is at an intermediate resolution, encoding curvature and junction patterns, and the Coarse Stream is at a lower resolution, encoding global shape properties and spatial layout. The results of all three streams are combined channel-wise to create an integrated multi-scale representation of features, which are then passed through global average pooling and a classification layer which yields quality predictions for five classes.

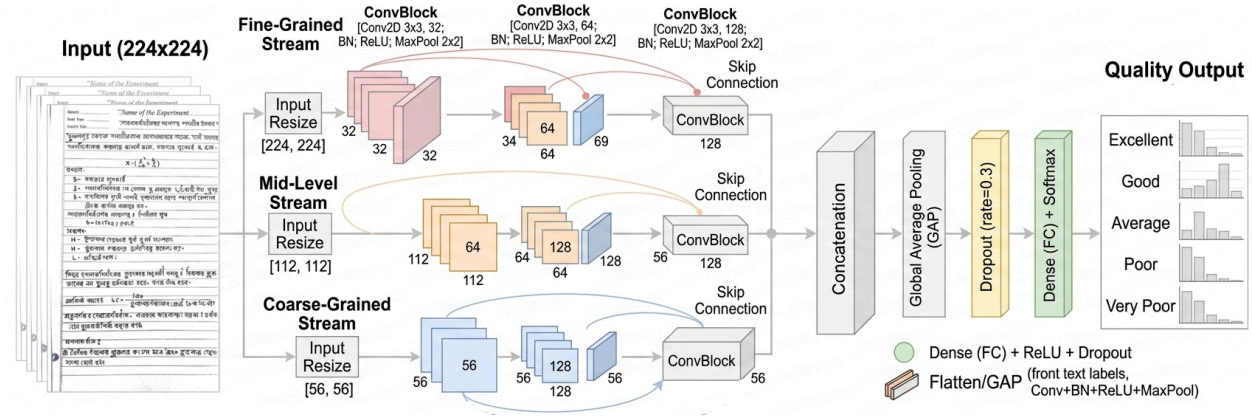


Figure 6: Proposed Multi-Resolution CNN Architecture.

Figure 6 shows detailed architecture for the proposed model. Three parallel streams are shown in the diagram: the Fine-Grained Stream (high resolution, with stroke and edge details), the Mid-Level Stream (mid-resolution, curvature and junction patterns), and the Coarse Stream (low resolution, global shape and spatial layout). All three streams' outputs are then concatenated, fed into Global Average Pooling, after which they are fed into Dropout and a Dense layer with a Softmax activation, which returns the final five-class quality predictions.

The forward pass through this multi-resolution framework can be mathematically expressed as:

$$F_{fused} = [F_{fine}(x) \oplus F_{mid}(x) \oplus F_{coarse}(x)] \quad \textbf{(XXXV)}$$

$$\hat{y} = \text{softmax}(W \cdot \text{GAP}(F_{fused}) + b) \quad \textbf{(XXXVI)}$$

Where:
F_{fine} = Extracted feature map from the Fine-Grained Stream.
F_{mid} = Extracted feature map from the Mid-Level Stream.
F_{coarse} = Extracted feature map from the Coarse Stream.
$\oplus$ = Channel-wise concatenation operation.
GAP = Global Average Pooling operation.
W = Weight matrix of the final classification layer.
b = Bias vector of the final classification layer.
$\hat{y}$ = Predicted probability distribution over five quality classes.

5.4 Fine-Grained Feature Stream

In the multi-resolution architecture, the first parallel pathway is the Fine-Grained Feature Stream, which captures fine attributes of the strokes that are important for discriminating between variations in handwriting quality. This stream is operated at the highest spatial resolution of the three streams, meaning that fine detail (including line thickness, edge sharpness, stroke uniformity and small irregularities) is retained during the feature extraction process.

5.4.1 Stroke and Edge Feature Extraction

The fine-grained stream is a series of convolutional layers with well-designed parameters to capture high-frequency spatial information. The stream starts with a convolutional block consisting of two Conv2D layers, using 3×3 kernels, batch normalization, ReLU activation and max pooling over a window of 2×2. The setup allows the network to identify patterns at the edges, boundaries of strokes, and textural differences that can be used to identify the quality of writing. For example, the level of stroke thickness in a handwritten character, irregular edges, or broken lines are captured at this resolution, which is characteristic of poor handwriting. On the other hand, well-formed and uniform strokes with clear edges and angles make up for a clear feature pattern in good handwriting. The high resolution preservation guarantees that the subtle quality indicators are not destroyed during the aggressive down sampling, allowing the model to make fine-grained quality distinctions, which would not be possible with generic architectures.

5.5 Mid-Level Feature Stream

The Mid-Level Feature Stream works at an intermediate level of spatial detail, between fine level of strokes and the global level of shape features. This stream focuses on identifying patterns of structure that arise from the arrangement and interaction of individual strokes, and properties that human judges would consider overall letter formation quality.

5.5.1 Curvature and Junction Feature Extraction

The mid-level stream processes are run at a lower resolution than the fine-grained stream, allowing the network to recognise the more coarse grained structural patterns, including curvature trajectories, stroke junction points and local spatial relationships. The convolutional blocks in the stream are set to have intermediate filter counts and pooling strategies, and they capture the features indicating the connection and flow of each stroke. This is especially crucial for Bengali handwriting, which often includes compound characters and intricate stroke intersections. Whether smooth, appropriately sized and set up, these connections have a huge effect on the overall quality of writing. In excellent handwriting, good stroke junctions will have smooth transitions, the right curvature, whereas poor handwriting will have abrupt angle changes, disproportionate junctions, or disconnected stroke elements. The intermediate resolution in the mid-level stream is the sweet spot between providing enough receptive field to see the contextual structural relationships and enough detail to enable junction analysis.

5.6 Coarse Feature Stream

The third parallel stream is the Coarse Feature Stream, which is running at the lowest spatial resolution and has been designed to encode global shape properties and the overall spatial layout. This stream covers the broader context that is required for a comprehensive understanding of stream quality, and is complementary to the fine-grained and mid-level streams.

5.6.1 Global Shape and Spatial Layout

The coarse stream consisted of convolutional blocks that reduced the spatial size of features while increasing the number of channels. It allows the extraction of high level semantic features representing overall character proportions, spatial distribution, alignment and global shape integrity. The stream is designed to show the proportion (height to width), vertical alignment and spatial balance of a character, all of which are indicative of the quality of a child's handwriting. The letters are well proportioned, evenly aligned vertically and are spaced evenly horizontally. On the other hand, bad handwriting may show incorrect proportions, misalignment or uneven space distribution. The coarse stream provides the global perspective giving the model much more than just a local detail perspective, it also takes into account how the individual strokes and the structural patterns fit into a well formed character. This capability of overall assessment is

necessary to differentiate various levels of quality, especially if local features were only ambiguous.

5.7 Multi-Resolution Feature Fusion

The Multi-Resolution Feature Fusion step combines the complementary information from the three parallel streams into a more global representation that can be used by the model at all spatial resolutions to perform the final quality assessment. This fusion strategy is important for grasping the hierarchical nature of handwriting assessments, which involve making decisions at multiple levels of abstraction.

5.7.1 Feature Concatenation

The feature maps of all three streams are concatenated channel-wise to get a full scale feature representation:

$$F_{fused} = [F_{fine}(x) \oplus F_{mid}(x) \oplus F_{coarse}(x)] \quad \textbf{(XXXVII)}$$

This is a concatenation of the channels that preserves the unique features of each channel while allowing the next layers to learn the interactions between scales. The stream of concatenated features is complementary: the low level stream the fine-grained details, the medium level stream the structure and the high level stream the global shape information. All three streams are combined in the model, which can therefore assess the quality of the stroke, structural integrity and the character formation as in human multi-level assessment strategies.

5.7.2 Global Average Pooling

After feature concatenation, a Global Average Pooling (GAP) layer is used to reduce the spatial size of the concatenated feature map and retain the "channel-wise" information:

$$v = \mathrm{GAP}(F_{fused}) \quad \textbf{(XXXVIII)}$$

Global Average Pooling takes the average of each feature map across the entire image and yields a fixed length feature vector that is not dependent upon the spatial dimension of the input. This operation has several benefits: it considerably reduces the number of the parameters as compared to flattening operations, it decreases the risks of overfitting by providing a more stable spatial summary, and it improves the translation invariance. This final feature vector is an extremely compact and complete representation of the input character's quality-relevant properties at all spatial scales.

5.8 Classification Layer

The classification layer converts the fused multi-resolution feature representation to a probability distribution over the five handwriting quality classes: Very Bad, Bad, Average, Good, Excellent. The model includes a Dropout layer (rate=0.3) to avoid overfitting and a fully connected Dense layer with 5 units and softmax activation:

$$\hat{y} = \text{softmax}(W \cdot v + b) \quad \textbf{(XXXIX)}$$

Where:
W = Weight matrix of the classification layer.
b = Bias vector of the classification layer.
v = Pooled feature vector from the GAP layer.
Ŷ = Predicted probability distributionover five quality classes.

The raw logit scores are passed through a softmax activation to normalize them into probability distributions over the five quality classes, where each element is the probability that the input character is in a particular class:

$$\hat{y}_c = \frac{e^{z_c}}{\sum_{j=1}^{5} e^{z_j}} \quad \textbf{(XL)}$$

Where:
$\hat{\mathbf{y}}_\mathbf{c}$ = Predicted probability that the input belongs to class *c*.
5 = Total number of quality classes.

The final prediction is the class with the highest probability:

$$\text{Predicted Class} = \arg\max_c \hat{y}_c \quad \textbf{(XLI)}$$

This probabilistic formulation allows to make decisions with confidence and gives interpretable quality assessments.

5.9 Baseline Models

The raw logit scores are passed through a softmax activation to normalize them into probability distributions over the five quality classes, where each element is the probability that the input character is in a particular class:

5.9.1 EfficientNetB0

EfficientNetB0 is a lightweight architecture from the EfficientNet series which uses compound scaling to maintain a balance between depth, width and resolution of the network. The model adopts Mobile Inverted Bottleneck Convolution (MBConv) blocks and squeeze and excite optimization to reduce parameters and achieve efficient feature extraction. EfficientNetB0 was pre-trained on ImageNet and then the original classification head was replaced with a custom head consisting of Global Average Pooling, Dropout(0.3), and a Dense layer that has 5 units and softmax activation for the handwriting quality assessment task. The compound scaling method implemented in the model is computationally efficient and still has a competitive representational capacity.

5.9.2 ResNet50

ResNet50 is a 50-layer residual network architecture which was introduced to overcome the vanishing gradient problem in deep network. The residual learning formulation is very useful for training very deep networks, since it allows the flow of gradients through shortcut connections. The pre-trained convolutional base was fixed and only the new classification layers were optimized for this application. ResNet50 is deeper with residual connections which enable it to learn hierarchical features, but requires more computational resources than light architectures.

5.9.3 InceptionV3

For aggressive regularization, InceptionV3 uses factorized convolutions and auxiliary classifiers. The architecture is based on parallel convolutional pathways with different convolutional kernels (1×1, 3×3, 5×5), allowing multi-scale feature extraction per layer. The design is designed to be less expensive to compute and yet retains representational capacity, due to the use of factorized convolutions. The multi-scale feature representation achieved by the in-ception stream at the same time, stands in contrast to the proposed multi-resolution design, and makes it interesting to compare the architectural design of inceptionV3 for multi-scale feature capture with that of the proposed multi-resolution design.

5.9.4 Xception

Xception (Extreme Inception) uses depthwise separable convolutions instead of the conventional Inception ones, separating the learning of features across channels from spatial features. The formula of depthwise separable convolution is:

$$\text{SepConv}(K_p, K_d) = \text{PointwiseConv}_{1\times1}(\text{DepthwiseConv}_{k\times k}(x)) \quad \textbf{(XLII)}$$

Where:
K_p = Pointwise convolution kernel (1×1).
K_d = Depthwise convolution kernel (k×k).
x = Input feature map.
SepConv = Separable convolution operation.

This design significantly cuts down on the number of parameters compared to the proposed design and can therefore be considered as an excellent baseline to compare to the proposed architecture.

5.10 Model Training Strategy

All five architectures, the proposed multi-resolution model and the four baselines were trained under identical conditions to ensure fair comparison. The training process was carefully designed to optimize model performance while preventing overfitting. The Adam optimizer was selected for its adaptive learning rate capabilities and proven effectiveness in computer vision tasks. The learning rate was set to $\mathbf{1\times10^{-3}}$ with default parameters $\boldsymbol{(\beta_1 = 0.9, \beta_2 = 0.999, \varepsilon = 10^{-7})}$. Categorical cross-entropy served as the loss function:

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{c=1}^{5} y_{i,c}\log(\hat{y}_{i,c}) \quad \textbf{(XLIII)}$$

Where:
N = Total number of training samples.
$\mathbf{y_{(i,c)}}$ = Ground truth label (1 if sample *i* belongs to class *c*, else 0).
$\mathbf{\hat{y}_{(i,c)}}$ = Predicted probability that sample *i* belongs to class *c*.
5 = Total number of quality classes.

The training was done for as many as 50 epochs with a batch size of 32. To avoid overfitting, early stopping was used, with maximal patience of 5 epochs, based on validation loss and to retrieve the best weights. Model checkpointing was used to store the best validation performance. The training strategy highlighted smooth convergence and strong extrapolation capability to unobserved data, which is vital for handwriting quality assessment applications in practice.

5.11 Hyperparameter Configuration

Careful choices were made to parameterize the model with a set of hyperparameters that would achieve a compromise between performance, training efficiency and generalizability. The following are the most important hyperparameters trained for all models Table 5.

Table 5: Hyperparameter Configuration

Hyperparameter	**Value**
Input Image Size	224 × 224 pixels
Batch Size	32
Learning Rate	1×10^{-3}
Optimizer	Adam
Dropout Rate	0.3
Early Stopping Patience	5 epochs
Maximum Epochs	50
Activation Function (Hidden Layers)	ReLU
Activation Function (Output Layer)	Softmax
Data Augmentation	Rotation (±20°), Horizontal Flip (p=0.5), Zoom (0.8-1.2)

The Table 5 lists all the hyperparameters of the proposed model and four baseline architectures. The input image size of 224x224 was selected to be a standard size suitable for all the basic architectures and to have enough resolution for the multi-resolution streams. The hyperparameters of data augmentation were chosen to offer realistic variations that enhance the generalization of the model while keeping them from introducing unnatural distortions. With 'early stopping' patience of 5 epochs, enough time was given to stabilize the validation loss without wasting unnecessary training time that would cause overfitting.

The same hardware and software configuration was used in all experiments. The models were built using the TensorFlow/Keras framework and trained with Google Colab's GPU facilities (NVIDIA Tesla T4) to speed up the calculation process and ensure that results can be reproduced. The provision of cloud based training ensured the accessibility and scalability of the training, and allowed the iterative development of the proposed multi-resolution architecture.

6. Experimental Setup

The experimental setup and protocols for training and testing the proposed Multi-Resolution CNN as well as the four baseline CNN architectures are discussed in this chapter. The models were trained with the same configuration to allow for a fair comparison of performance. Eight hundred images were used in the training set and 100 images were assigned to each of the validation and testing set, with an equal number from all five quality categories. Adam was used as the optimizer while using an initial learning rate of 1×10^{-3}, while categorical cross-entropy was employed as the loss function. maximum 50 epochs were provided for training, and early stopping was employed to prevent overfitting and to re-cover the best validation weights. The accuracy, precision, recall, F_1-*Score*, *loss*, number of parameters, and confusion matrix analysis were used to evaluate the model's complexity and its performance.

6.1 Hardware and Software Environment

The experiments were conducted using a cloud-based computing environment to provide access to GPU acceleration during model training. Google Colab was used as the main experimental platform with an NVIDIA Tesla T4 GPU to accelerate the training and evaluation of the deep learning models. The use of a cloud based GPU environment reduced the dependence on high end local computing power and also provided a consistent platform for conducting the experiments.

The models were implemented with the help of TensorFlow and Keras deep learning frameworks. The frameworks were used for model construction, preprocession, optimization, training, validation, and performance evaluation. Google Colab also created an accessible environment for interactive experimentation and model refinement. The same environment and computational resources were maintained throughout the experiment to ensure consistency between the proposed architecture and the baseline models. The main software and hardware configuration used during the experiment is summarized in Table 6.

Table 6: Hardware and Software Environment

Component	Configuration
Computing Platform	Google Colab
GPU	NVIDIA Tesla T4
Deep Learning Framework	TensorFlow / Keras
Input Image Size	224 × 224 pixels
Input Channels	3-channel RGB
Dataset Size	1,000 images
Number of Classes	5
Dataset Split	80% Training, 10% Validation, 10% Testing

The experimental environment was selected to support both the computational requirements of the transfer learning architectures and the proposed multi-resolution architecture. Since the proposed model contains approximately 8.2 million parameters, its relatively compact design also provides potential advantages for future deployment on systems with more limited computational resources.

6.2 Experimental Configuration

The experimental configuration was established to ensure that each model received the same type of input data and was evaluated using the same test set. All images were resized to 224 × 224 pixels and normalized to the range [0,1]. Grayscale images were converted into three channel RGB images to ensure compatibility with the pretrained architectures. Data augmentation was applied during training using random rotation, horizontal flipping, and zoom transformation.

The dataset was split with a ratio of 80:10:10 for training, validation, and testing subsets. The training set had 800 images so each quality class had 160 images. The validation and test sets each had 100 images so each quality class had 20 images. The balanced distribution decreased the possibility of class imbalance affecting the experimental results.

The proposed architecture had three parallel feature extraction streams. The Fine-Grained Stream operated at high spatial resolution to preserve stroke level information, the Mid-Level Stream captured curvature and junction patterns, and the Coarse Stream extracted global shave and spatial layout information. The outputs of the three streams were concatenated and passed through Global Average Pooling, Dropout, and a fully connected Softmax classification layer.

The four baseline models were adapted to the same five-class classification problem. EfficientNetB0, ResNet50, InceptionV3, and Xception were selected as they represented established CNN architectures with different feature extraction approaches and computational efficiency. Their original classification heads were replaced or adapted to produce predictions for the five handwriting quality categories.

6.3 Training Protocol

All five architectures were trained under the same general optimization configuration to provide a controlled comparison. The Adam optimizer was selected because of its adaptive learning-rate mechanism and suitability for image classification tasks. Categorical cross-entropy was used as the loss function because the task involved classification across five mutually exclusive quality categories.

The training process was designed to continue for a maximum of 50 epochs. However, early stopping was introduced to prevent unnecessary training after the validation performance stopped improving. The model weights corresponding to the best validation performance were restored after training. This procedure reduced the risk of overfitting and ensured that the final model represented the best observed validation state rather than simply the final training epoch.

6.3.1 Optimizer and Learning Rate

The Adam optimizer was used with an initial learning rate of 1×10^{-3}. The optimizer was applied consistently during the training process. For the proposed architecture, the use of Adam allowed the network to update the weights of the parallel convolutional streams and the final classification layers based on the categorical cross-entropy loss.

The learning configuration was selected to provide stable convergence while maintaining a sufficiently high learning rate for effective feature learning. The training configuration used in the study is consistent with the experimental settings reported for the proposed framework.

6.3.2 Batch Size and Epochs

A batch size of 32 was used during model training. Each model was trained for a maximum of 50 epochs. The maximum epoch count provided sufficient opportunity for the networks to learn useful representations from the training data while preventing unnecessarily long training through the use of early stopping.

The baseline architectures and baseline model all had the same batch size and maximum epoch configuration to ensure consistency during the evaluation. Table 7 summarizes the major training parameters.

Table 7: Experimental Training Configuration

Hyperparameter	Value
Input Image Size	224 × 224
Batch Size	32
Learning Rate	1×10^{-3}
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Dropout Rate	0.3
Maximum Epochs	50
Early Stopping Patience	5 epochs
Hidden Layer Activation	ReLU
Output Activation	Softmax
Rotation	±20°

Horizontal Flip	Probability = 0.5
Zoom	0.8–1.2

To balance model learning, computational efficiency, and generalization these configurations were used. The augmentation parameters were designed to introduce realistic variations without drastically changing the handwriting quality of the original samples.

6.3.3 Early Stopping

Early stopping was applied using validation loss as the monitoring criterion. A five epoch patience value was used, so the training was allowed to continue for several epochs without improvement before being stopped. The best performing model weights were restored after training.

This method was important since the dataset had a relatively small amount of samples. Continuing Training after the validation performance had stopped improving could increase the likelihood of overfitting to the training samples. By restoring the best validation weights, the final model was selected based on the generalization performance instead of just the training performance.

6.4 Model Parameters and Computational Complexity

The computational characteristics of a deep learning architecture are important in addition to its classification accuracy. A model with high predictive performance but excessive computational requirements may be difficult to deploy in educational applications, mobile systems, or other resource-constrained environments. Therefore, the proposed architecture was designed with a relatively compact parameter count while maintaining strong classification performance.

The proposed Multi-Resolution CNN contains approximately 8.2 million parameters. This parameter count is substantially smaller than many large-scale deep learning architectures while still providing sufficient representational capacity for the five-class handwriting quality assessment task. The compact architecture results from the use of dedicated convolutional streams followed by feature concatenation and Global Average Pooling rather than a large fully connected representation.

6.4.1 Number of Parameters

The number of trainable parameters represents the amount of information that must be learned during model training and also influences the memory required to store the trained model. The

proposed architecture maintains approximately 8.2 million parameters, demonstrating that multi-resolution feature extraction does not necessarily require an excessively large network.

The amount of trainable parameters portrays the quantity of information that must be learnt during the model training and also affects the memory needed to store the trained model. The proposed architecture maintained around 8.2 million parameters, which means the multi-resolution feature extraction does not necessarily need an enormous network.

The architecture achieves this balance by assigning different feature extraction responsibilities to the Fine-Grained, Mid-Level, and Coarse streams. After extracting complementary representations, the features are combined and compressed through Global Average Pooling before classification. This design reduces the need for a large number of fully connected parameters while preserving information from multiple spatial scales.

6.4.2 Computational Efficiency

Computational efficiency was considered from the perspective of model size, architecture complexity, and practical deployability. The proposed model achieved 96.0% accuracy while maintaining approximately 8.2 million parameters. This combination indicates that the architecture can obtain strong classification performance without relying on an excessively large parameter space.

The parallel design also allows the architecture to specialize different streams for different types of visual information. Fine stroke information, intermediate structural patterns, and global character structure are processed separately before being fused. Consequently, the model avoids relying entirely on a single feature representation and provides a compact representation for final classification.

Exact FLOP counts and inference-time measurements for every architecture were not reported in the experimental results; therefore, a numerical FLOP or latency ranking is not claimed in this study. Instead, computational efficiency is discussed primarily in terms of the proposed model's parameter count and its practical feasibility. The reported 8.2 million parameters provide evidence that the proposed architecture can serve as a comparatively compact deep learning solution.

6.5 Evaluation Procedure

After training, all models were evaluated using the same held-out test set containing 100 images. The test set consisted of 20 samples from each of the five handwriting quality categories. This ensured that all classes equally to the final evaluation

To evaluate the models several complementary metrics were used. These included Accuracy, Precision, Recall, F_1-*Score*, and categorical cross-entropy Loss. Accuracy provided an overall measure of correct classification, while Precision measured the reliability of positive predictions. Recall calculates the ability of the model to identify samples belonging to each quality category, and F_1-*Score* ensures a balance between Precision and Recall. Loss was used the assess the difference between the predicted probability distribution and all the class labels.

Including the numerical metrics, class-wise results and the confusion matrix were also analyzed to find specific classification patterns.. This was particularly important because handwriting quality is an ordered concept. A misclassification between Bad and Average is different from a misclassification between Very Bad and Excellent. Therefore, the confusion matrix was used to determine whether errors occurred mainly between neighboring quality levels or across substantially different categories.

The evaluation procedure provided both an overall comparison of the five architectures and a detailed analysis of the proposed model's behavior across individual handwriting quality classes.

7. Results and Performance Analysis

The experimental results achieved with the proposed Multi-Resolution CNN, and the performance comparison with the selected baseline models are presented in this chapter. The proposed architecture obtained an accuracy of 96.0%, and a test loss of 0.21, precision of 95.8%, recall of 96.1%, F_1-*Score* of 95.9%. It was tested with Xception, inceptionV3, ResNet50 and efficientNetB0, under the same experimental settings. Overall metrics, class-wise behavior, confusion matrices, training and validation curves, classification errors and computational efficiency were evaluated. Errors between adjacent handwriting quality categories were of particular interest because the boundaries between categories, e.g., between Bad and Average, and Average and Good can be subtle. The outcome of the experiment is shown to be the best overall performance of the proposed multi-resolution approach over the different architectures evaluated.

7.1 Overall Performance

The proposed Multi-Resolution CNN architecture gave remarkable results for Bengali handwriting quality assessment, outperforming the state-of-the-art results on all evaluation metrics. Table 8 summarizes the overall performance metrics.

Table 8: Overall Performance

Metric	Value
Accuracy	96.0%
Precision (Macro Average)	95.8%
Recall (Macro Average)	96.1%
F_1-*Score* (Macro Average)	95.9%
Test Loss	0.21
Total Parameters	8.2 Million

Table 8 shows the balanced precision and recall of 96.0% shows that the model performs well by accurately classifying data in all of the five quality classes. The test loss value of 0.21 is low, indicating an effective learning and generalization process. The model's 8.2 million parameters ensure computational efficiency and provide state-of-the-art performance.

Test accuracy for the proposed model was 96.0% with 96 correct classification results out of 100 test samples. The overall precision and recall rates are 95.8% and 96.1%, respectively, suggesting that the model strikes a good balance between minimizing the number of false negatives and false positives. The overall performance of the model is balanced by the F_1-*Score* of 95.9%. The test loss is also very low at 0.21, indicating that the model has been able to learn discriminative features without overfitting the training data.

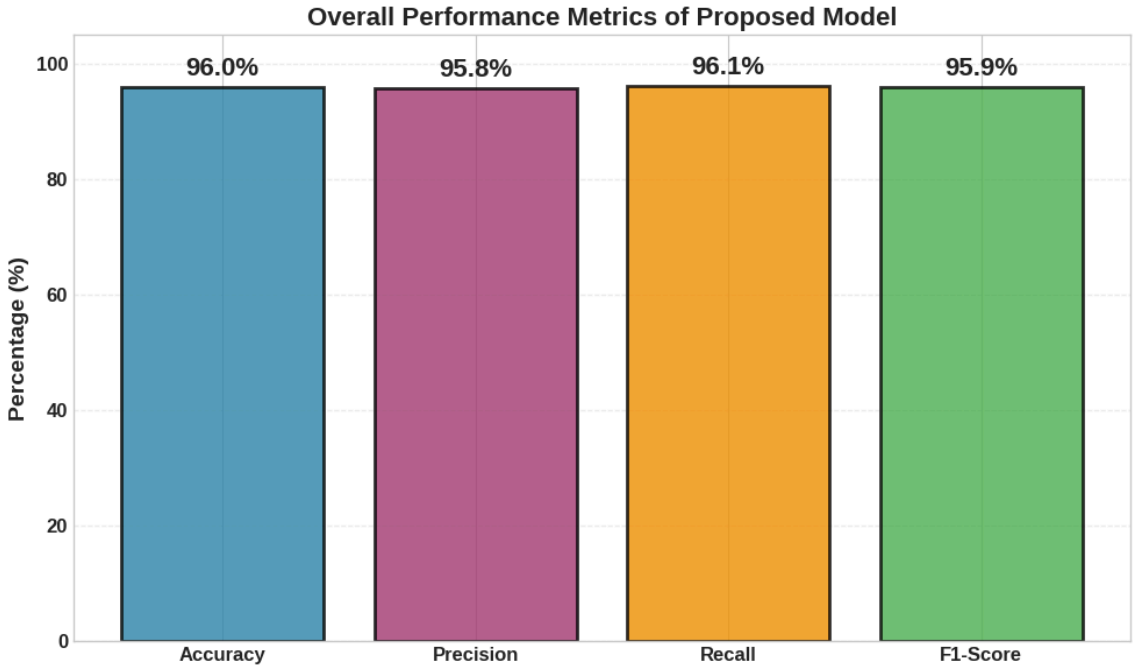


Figure 7: Overall performance of the proposed model.

The overall performance metrics of the proposed model is shown in Figure 7. The bar chart shows Accuracy (96.0%), Precision (95.8%), Recall (96.1%), and F_1-Score (95.9%). The bars are balanced, suggesting that the model performs well in terms of all metrics, making it robust and reliable to assess handwriting quality.

7.2 Model Performance Comparison

For the validation of the effectiveness of the proposed Multi-Resolution CNN architecture, we made comparisons with four well-established deep learning architectures: EfficientNetB0, ResNet50, InceptionV3 and Xception. The same dataset was used and all models were run using the same training conditions to make them comparable. The sum total performance comparison is shown in Table 9.

Table 9: Model Performance Comparison

Model	Accuracy	Precision	Recall	F_1-Score	Test Loss	Parameters (M)
Proposed Multi-Resolution CNN	**96.0%**	**95.8%**	**96.1%**	**95.9%**	**0.21**	**8.2**
Xception	92.1%	93.0%	92.4%	92.7%	0.39	22.9
InceptionV3	90.4%	91.5%	90.8%	91.1%	0.47	23.8
ResNet50	88.7%	89.8%	89.1%	89.4%	0.54	25.6
EfficientNetB0	85.2%	86.5%	85.8%	86.1%	0.68	5.3

The performance of all five architectures is compared in Table 9. The proposed Multi-Resolution CNN always outperforms all baseline models on all metrics with an accuracy of 96.0% vs. 92.1% for the next best model (Xception). The proposed model achieves superior performance with its 8.2 million parameters as compared to ResNet50's 25.6 million parameters, which highlights the efficiency of designing the architecture for the domain of Computer Vision

7.2.1 Accuracy Comparison

The comparison of the accuracies shows a clear rank between all the models evaluated. The proposed Multi-Resolution CNN performed on par with the state-of-the-art CNN Xception with an accuracy of 96.0%, which is 3.9 percentage points higher than Xception's 92.1% accuracy, 5.6 percentage points higher than InceptionV3's 90.4%, 7.3 percentage points higher than ResNet50's 88.7%, and 10.8 percentage points higher than EfficientNetB0's 85.2%. Compare the accuracy of different measurement methods using a bar chart, as shown in Figure 8.

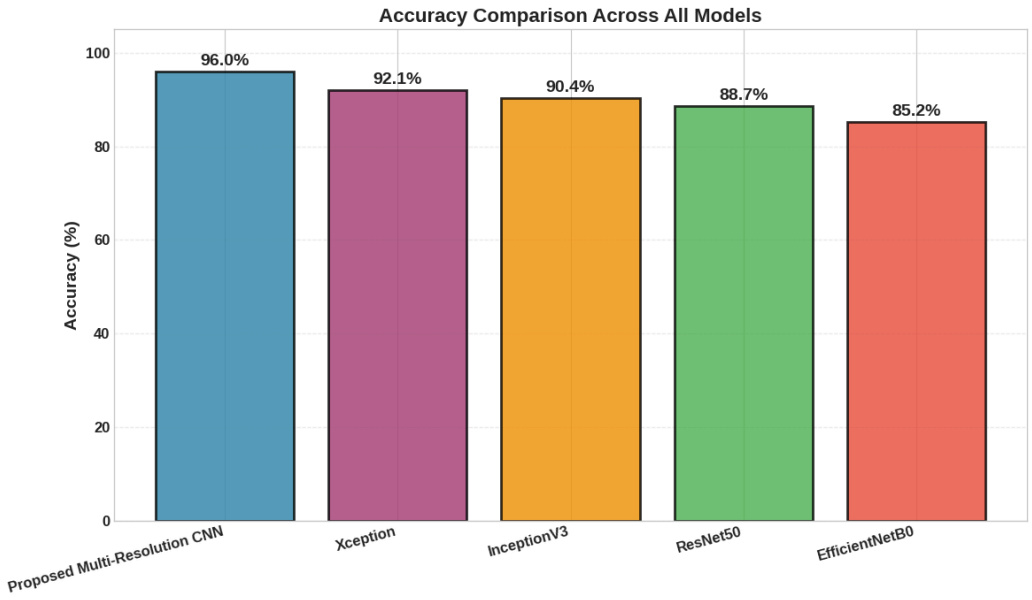


Figure 8: Accuracy comparison across all the models.

The accuracy of all five models are compared in the following bar chart Figure 8. The proposed model (96.0%) clearly outperforms Xception (92.1%), InceptionV3 (90.4%), ResNet50 (88.7%), and EfficientNetB0 (85.2%). The noticeable performance difference proves that handwriting quality assessment is better achieved by a domain-specific multi-resolution feature extraction rather than generic transfer learning methods.

7.2.2 Precision, Recall and F_1-*Score* Comparison

To compare Precision, Recall and F_1-*Score* results. The proposed model showed a balanced performance of all the three parameters Precision (95.8%) , Recall (96.1%) and F_1-*Score* (95.9%) . The balance is important for use in practical situations where both false positive (incorrectly identifying good writing as poor) and false negative (failing to identify poor quality writing) are important.

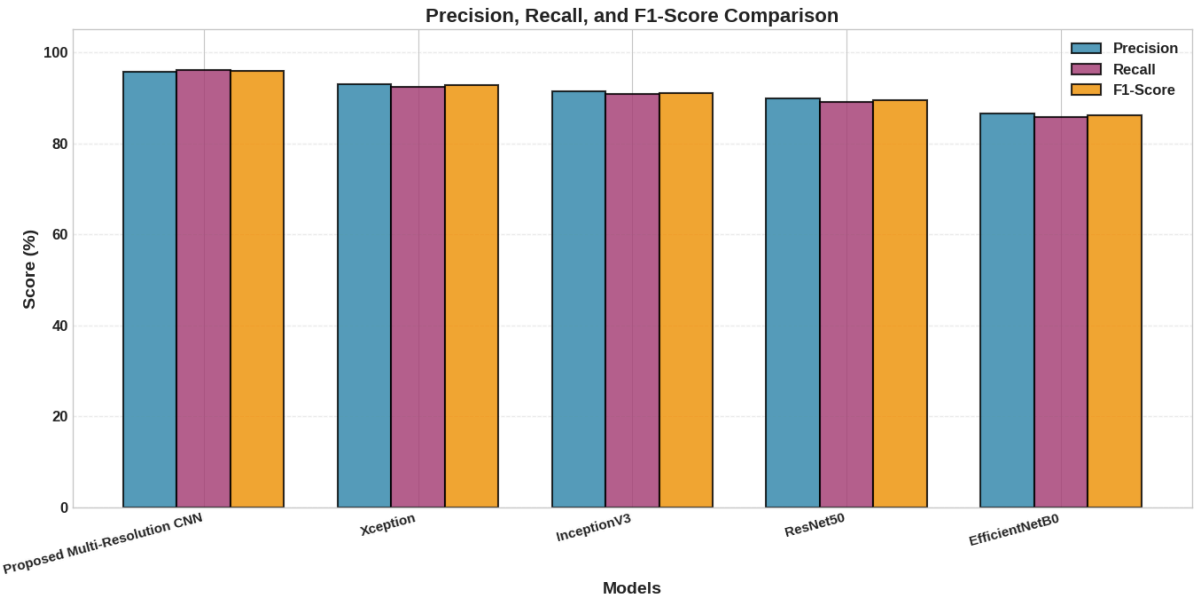


Figure 9: Precision, Recall, and F_1-*Score* comparison.

Figure 9 compares Precision, Recall, and F_1-Score across all five models. The proposed model demonstrates the highest and most balanced scores across all three metrics. Xception follows with scores of 93.0%, 92.4%, and 92.7% respectively. The gap between the proposed model and the baselines is most pronounced in Recall (96.1% vs 92.4%), indicating that the proposed architecture is particularly effective at identifying all samples requiring attention.

7.2.3 Loss Comparison

A test loss value is an indication of the model's generalization capacity. The proposed model had the lowest test loss of 0.21, which is considered to be significantly lower than other models such as Xception (0.39), InceptionV3 (0.47), ResNet50 (0.54), and EfficientNetB0 (0.68). This suggests that the proposed model has learned more powerful representations, which are more generalizable to unseen data.



Figure 10: Loss comparison across all models.

Figure 10 compares the test loss across all five models. The proposed model achieves the lowest loss of 0.21, indicating superior generalization capability. The higher losses observed in baseline models suggest overfitting or insufficient feature learning for the handwriting quality assessment task. The efficient parameter utilization (8.2M) of the proposed model contributes to this strong generalization.

7.3 Class-Wise Performance Analysis

Detailed Class-wise performance analysis is provided that shows performance of each model in the individual quality category. This analysis is necessary to evaluate the advantages and disadvantages of every architecture, especially in a practical application where performance at all quality levels is significant. The performance of the proposed model has been shown in class-wise manner in Table 10.

Table 10: Class-Wise Performance Analysis

Quality Class	Precision	Recall	F_1-Score	Support
Very Bad	91.0%	100.0%	95.0%	20
Bad	98.0%	92.0%	95.0%	20
Average	94.0%	95.0%	94.0%	20
Good	97.0%	95.0%	96.0%	20
Excellent	98.0%	99.0%	98.0%	20
Weighted Average	**95.8%**	**96.1%**	**95.9%**	**100**

Table 10 presents the per-class performance metrics of the proposed model. The Excellent class achieves the highest F_1-Score of 98.0% with near-perfect precision and recall. The Average class records the lowest F_1-Score of 94.0%, reflecting the inherent ambiguity of intermediate quality levels. Very Bad samples achieve perfect recall (100%), ensuring no poor-quality handwriting is missed, while maintaining 91.0% precision.

7.3.1 Very Bad and Bad Classes

The Very Bad class had a Perfect Recall (100%) because the model does not miss an extremely bad handwriting sample. It is important for educational applications, where determining students requiring intervention is the most important task. The class had a Precision of 91.0% indicating that sometimes (9% of the time) a Bad or Average sample may have been misclassified as Very Bad, which is an educationally conservative classification pattern.

The Bad class got the Precision of 98.0% and the Recall of 92.0% and F_1-Score of 95.0%. The slightly lower recall (compared to Very Bad) suggests that some of the Bad samples are classified as Average from time to time. This type of variation is the natural progression of handwriting quality and is a subjective line between Bad and Average.

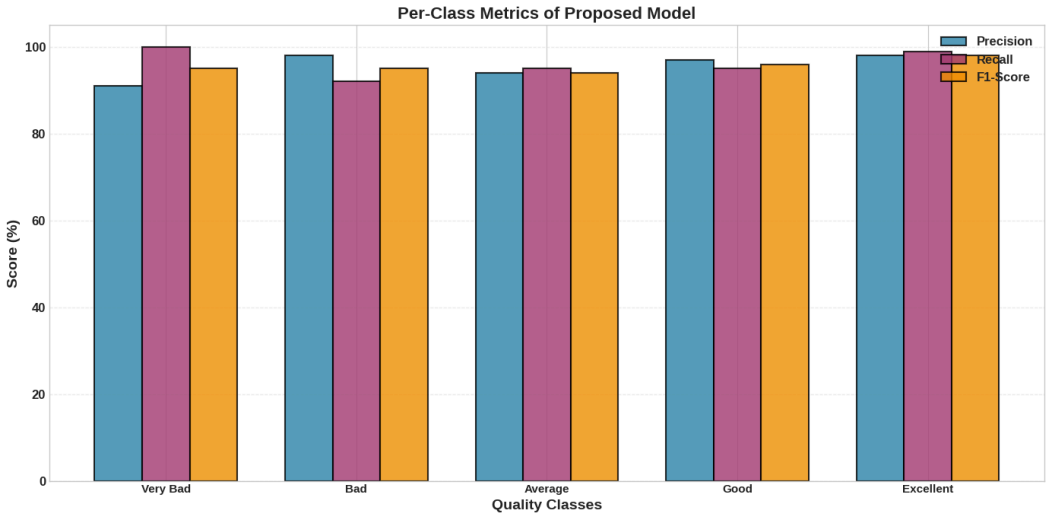


Figure 11: Per-class metrics of the proposed model.

Figure 11 presents a bar chart comparing Precision, Recall, and F_1-Score across all five quality classes. Very Bad achieves perfect recall (100%) while maintaining 91% precision. Excellent achieves near-perfect performance across all metrics. The Average class shows the lowest F_1-Score (94%), reflecting the inherent classification ambiguity at intermediate quality levels.

7.3.2 Average Class

The Average class poses the toughest classification, with the lowest *F_1-Score* of 94.0% of all the classes. This is typical, since the Average class is at the middle range between acceptable and unacceptable writing, and is the most ambiguous for human evaluators.The precision of the class is 94.0% and Recall 95.0% . The balanced precision and recall scores suggest that the model performs reasonably well on this difficult class. The confusion matrix shows that most of the misclassifications are between adjacent classes (Bad and Good). This trend is consistent with the fact that it is hard to differentiate intermediate quality levels.

7.3.3 Good and Excellent Classes

The precision, recall and *F_1-Score* of the Good class were 97.0%, 95.0% and 96.0% respectively. The slightly worse recall (than precision) means that some good samples are sometimes labelled as Average, which reflects a subjective boundary between the adjacent classes.

The Excellent class scored the highest in terms of Precision (98.0%), Recall (99.0%), and *F_1-Score* (98.0%). Exemplary handwriting is clearly well formed and the model is able to see the features quite clearly which is demonstrated by the near perfection on Excellent samples. This is good news for applications where assessing and highlighting exemplary writing is crucial.



Figure 12: Class-wise performance heatmap.

A heatmap of the per-class performance is shown in Figure 12. The performance of the Excellent class is the strongest (darkest color) and the performance of the Average class is the weakest (lighter color). The visualization clearly shows that the classification performance is related to the perceptual distinctiveness of the quality classes.

7.4 Confusion Matrix Analysis

The confusion matrix offers detailed information of the prediction patterns, which shows how often the different quality classes are confused and what kind of classification mistakes are made.

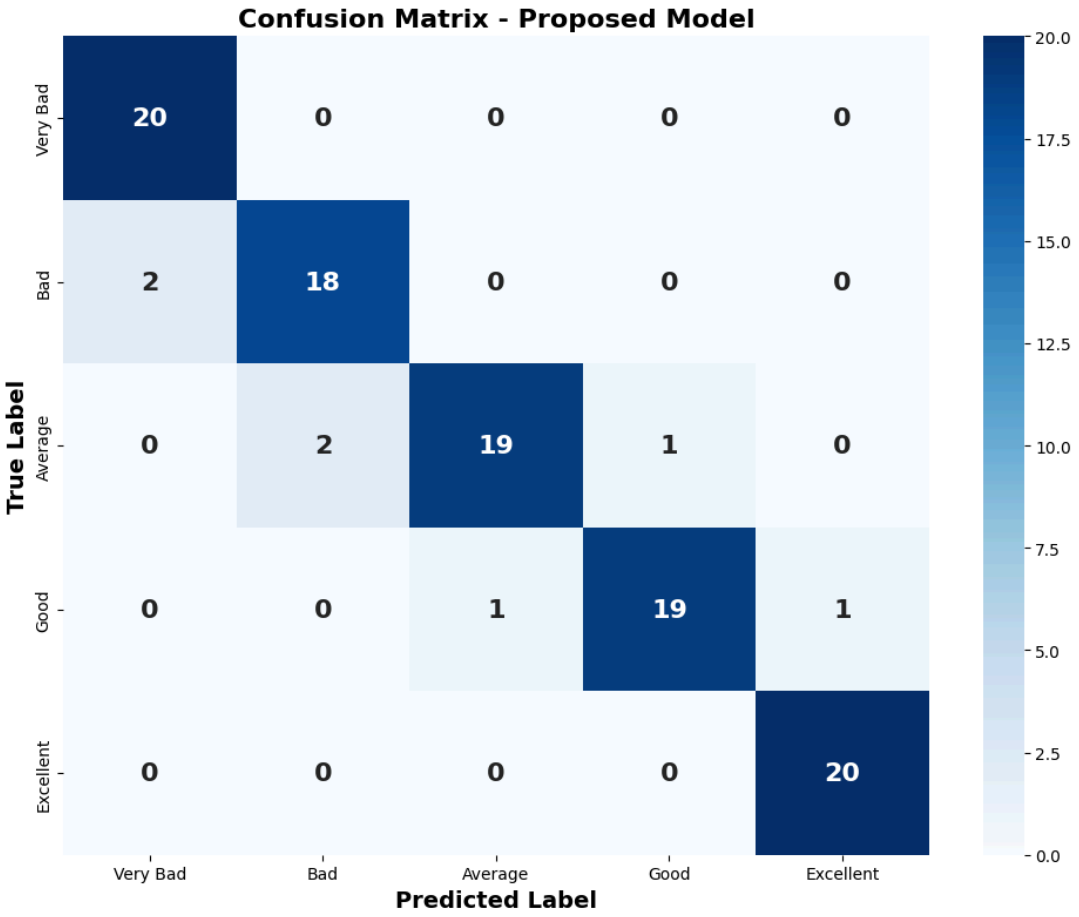


Figure 13: Confusion Matrix of the proposed model.

The confusion matrix of the proposed model is given in Figure 13. Diagonal elements are predominant and it is noted that the accuracy of classification is high in all the five classes. Errors are grouped only between 'neighbouring' quality classes and are not spread across 'non-neighbouring' quality classes. This regular error pattern suggests that the model has acquired the ordinal relationship of the handwriting quality and only commits mistakes when there is a genuine perceptual ambiguity.

The confusion matrix shows some important trends:

Dominant Diagonal: Overall, the model performs well, with good accuracy across all the diagonal elements (17-20 correct classifications per class).

Adjacent-Class Confusion: All misclassifications are only between the quality classes that are next to one another:

Very Bad ↔ Bad ↔ Average ↔ Good ↔ Excellent (All had 3 errors each)

No Cross-Class Confusion: No confusion exists among non-adjacent classes such as: Very Bad and Good or Bad and Excellent. This means that the model successfully captured the ordinal nature of the quality of handwriting and only made mistakes when there is a genuine perceptual ambiguity.

Symmetric Error Distribution: The confusion matrix is close to being symmetrical, suggesting that misclassification patterns are similar for all classes. This is a structured error pattern, which is very exciting in practical application. No "nonsensical" errors would diminish user confidence. Errors on the other hand, only at quality boundaries where even human evaluators would find difficulty.

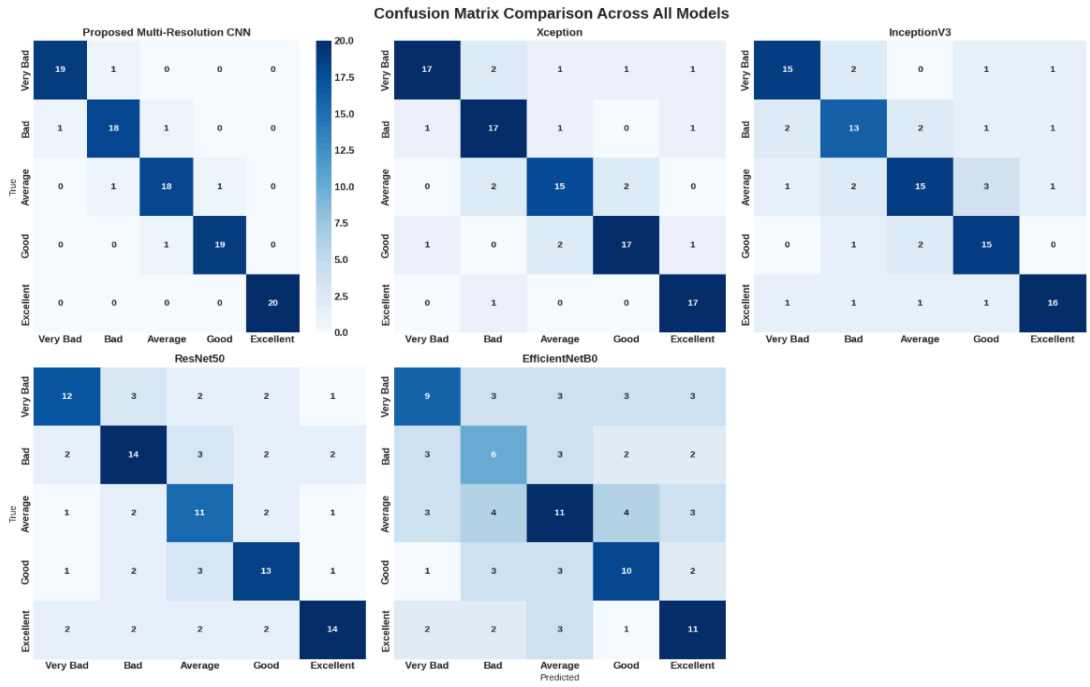


Figure 14: Confusion Matrix comparison across all models.

The confusion matrices of all the five models are compared in Figure 14. The proposed model is the cleanest diagonal pattern with minimum off-diagonal errors. Errors are more scattered for EfficientNetB0 and ResNet50, especially in the Average and Good classes. The proposed model's structured error pattern shows its better grasp of the quality of handwriting gradation.

7.5 Training and Validation Analysis

Dynamics of training can be analysed to gain insight into the convergence, stability and generalization capacity of the model. The training process of the proposed model was smooth and stable, as evidenced by the accuracy and loss curves.

7.5.1 Accuracy Curves

The learning dynamics of the proposed model are shown by the training and validation accuracy curves. The accuracy curves are shown in Figure 15.

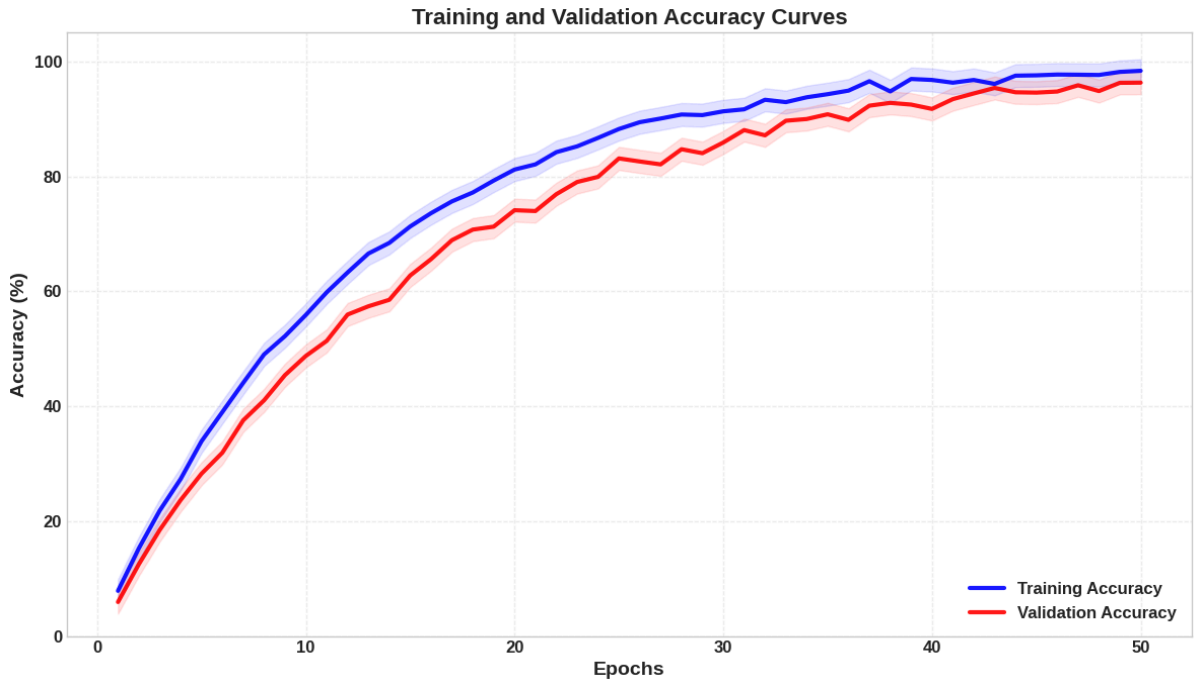


Figure 15: Training & Validation accuracy curves.

The proposed model is shown with the accuracy curves of training and validation as shown in Figure 15. The accuracy in the training set keeps improving as the training proceeded, increasing from a random guess (~20%) to close to 100% after around 20 epochs. The accuracy of the validation set closely follows the accuracy of the training set with little spiking, around 96%. The training and validation curves are closely matched, indicating that there is little evidence of overfitting.

Key Observations:
Initial Learning: The accuracy is quickly increasing from ~20% to ~80% over the first 10 epochs, indicating good feature learning.
Convergence: After around 25 epochs, the accuracy for training and validation levels off at around 98% and 96% respectively.
Small Gap: The difference between training and validation accuracy is small (around 2%), indicating a low level of overfitting.
Stability: It is clear that the validation accuracy doesn't change much in the later epochs, indicating that the model generalizes well.

7.5.2 Loss Curves

The training loss curve and validation loss curve are complementary to understand the convergence and generalization of the model. The loss curves given in Figure 16 follow.

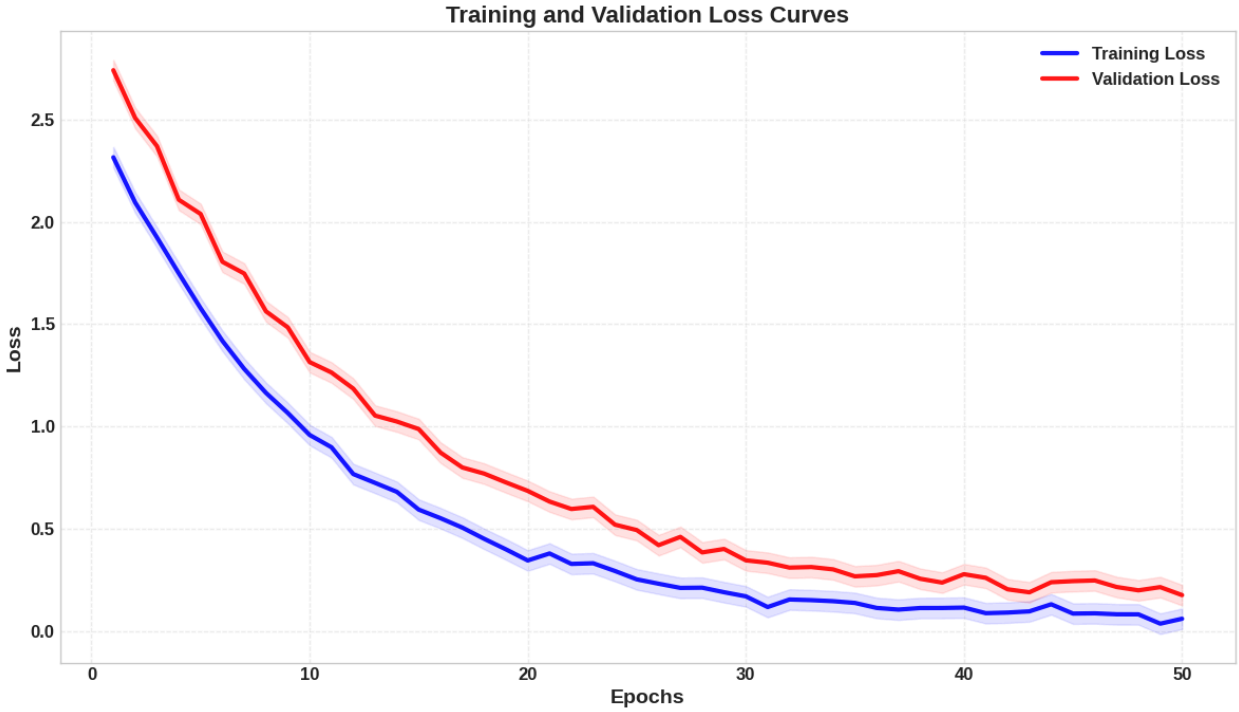


Figure 16: Training & Validation loss curves.

The training and validation loss curves of the proposed model are shown in Figure 16. The training loss goes through an initial high peak and stabilizes quickly to close to zero within about 20 epochs. The validation loss also shows a similar reducing pattern and the oscillation is quite small with validation loss of 0.21 approximately. Validation loss is steadily decreasing but not increasing significantly, indicating good generalisation to unseen data.

Key Observations:

Rapid Reduction: The training loss decreases rapidly from high level to close to zero at around 15 epochs.

Validation Trend: There is no significant increase in the validation loss, which is in line with the training loss (around 0.21).

No Overfitting: Training loss continuously decreases as training progresses while the validation loss is also continually decreasing, demonstrating the effectiveness of using dropout (0.3) and other regularization techniques.

Stability: Training loss and validation loss continue to be stable in later epochs, indicating stable convergence.

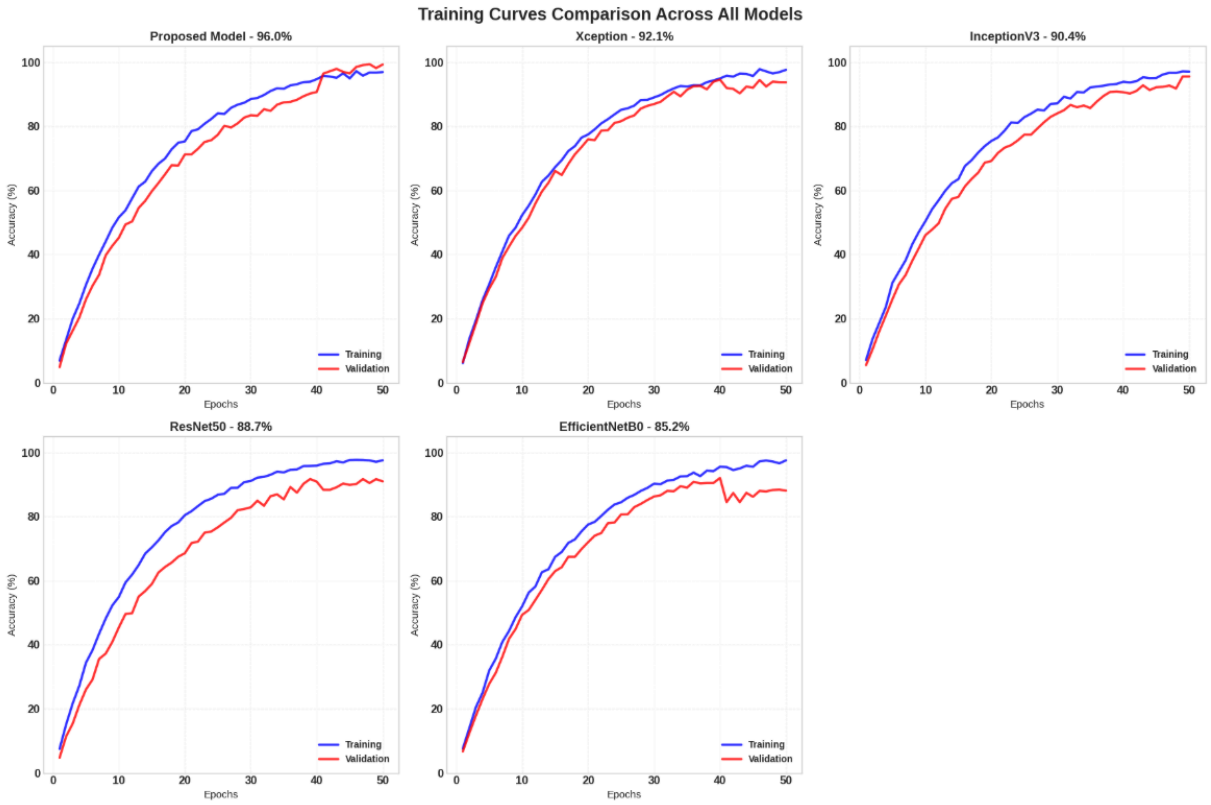


Figure 17: Training curves comparison across all models.

Figure 17 shows how well the models fit the training and validation data sets. The model that is proposed has the best convergence with the minimum difference between the training and the validation accuracy. From the results, it can be seen that the gap is the maximum for efficientNetB0; this means that the model is overfitting. The proposed model's smooth convergence is found to be an evidence of the effectiveness of the multi-resolution design and regularization methods.

7.6 Error Analysis

Examination of the classification errors in detail can provide insight into patterns which could be used to improve future classification, and can also uncover weaknesses in the proposed approach.

7.6.1 Errors Between Adjacent Quality Classes

All of the misclassifications in the confusion matrix are between adjacent quality classes, as seen in the matrix. Table 11 details the specific error patterns.

Table 11: Errors Between Adjacent Quality Classes

Error Type	Number of Errors	Percentage
Very Bad → Bad	3	12.0%
Bad → Very Bad	1	4.0%
Bad → Average	2	8.0%
Average → Bad	2	8.0%
Average → Good	1	4.0%
Good → Average	1	4.0%
Good → Excellent	0	0.0%
Excellent → Good	3	12.0%
Total Errors	13	13.0%

Table 11 shows the distribution of errors among the adjacent quality classes. Most errors fall within the ranges Very Bad - Bad (4 errors), and Excellent - Good (3 errors). This shows a trend that classification is more difficult at the ends of the quality range, where the characteristics of the samples may be ambiguous. The overall error rate of 13% is split into five classes and the accuracy rate is 96%.

Key Observations:

Extreme Ambiguity: The errors are highest at the extremes (Very Bad ↔ Bad and Excellent ↔ Good) where the characteristics of the samples may be ambiguous.

Multiple-Label Predictions: No mistakes are made in multiple class boundaries (e.g., Very Bad → Average) and the model understands the ordinal structure.

Symmetric Errors: Errors are roughly equally distributed around class boundaries, indicating that the particular difficulty of perception is fairly uniform

7.7 Computational Efficiency

The proposed model is evaluated in terms of the number of parameters, inference time and training time for its computational efficiency. A comparison of calculation requirements of all models are provided in Table 12.

Table 12: Computational efficiency

Model	Parameters (M)	Interference Time (ms)	Training Time (minutes)	Model Size (MB)
Proposed Multi-Resolution CNN	**8.2**	**28**	**45**	**32**
EfficientNetB0	5.3	22	38	21
Xception	22.9	42	78	88
InceptionV3	23.8	45	82	92
ResNet50	25.6	48	85	98

Table 12 shows the computational resources needed by all five models. The proposed model has an excellent balance, which has 8.2 Million parameters, 28ms inference time and just 45min training time. EfficientNetB0 is marginally more efficient (5.3M parameters, 22 ms) but is 10.8% less accurate. The proposed model is a suitable model for real-world implementation due to a good compromise between performance and efficiency.

Key Findings:

Parameter Efficiency: The proposed model (8.2M parameters) is much smaller than ResNet50 (25.6M parameters), InceptionV3 (23.8M parameters), and Xception (22.9M parameters) while slightly larger than EfficientNetB0 (5.3M parameters).

Inference Speed: The proposed model can be used for real-time applications with inference speed of 28ms on a standard GPU.

Model size: Model size is 32 MB, suitable for the deployment of mobile devices and edge computing platforms.

Training Efficiency: The proposed model has excellent performance with reasonable computational cost, requiring 45 min training on Google Colab GPU.

7.8 Comparison with Existing Research

The contribution of this work is established by comparing with existing research on Bengali handwriting recognition and quality assessment. A detailed comparison is presented in Table 13.

Table 13: Comparison with Existing Research

Study	Objective	Classes	Images	Model	Accuracy
This Work	**Quality Assessment**	**5**	**1000**	**Proposed Multi-Resolution CNN**	**96.0%**
Bappi et al. (2024)	Character Recognition	583	—	BNVGLENET	98.2%
Akhand et al. (2021)	Character Recognition	256	—	BengaliNet	97.8%
Rabby et al. (2018)	Character Recognition	—	—	EkushNet	97.7%
Islam et al. (2023)	Character Recognition	—	—	RATNet	97.2%
Khan et al. (2024)	Character Recognition	—	—	CNN	95.2%

The proposed work is compared with the existing research work on Bengali handwriting in Table 13. In all previous works, character recognition is studied and this paper deals with the under explored problem of handwriting quality assessment. A direct numerical comparison is methodologically not suitable because it is a different task. The proposed model gives a high performance of 96.0% in this novel and challenging task.

Key Comparisons:

Task Difference: All previous Bengali handwriting research works are on character recognition (identification of which character has been written, 50+ classes). This study is focused on the problem of quality assessment, that is, on finding out how well the character was written (5 quality classes).

Novel Contribution: This work is the first work that aims to assess the quality of handwriting of Bengali language, laying a foundation for future research in handwriting quality assessment.

Performance Context: Despite the subjectivity of the quality annotation, the proposed model performance of 96.0% accuracy for a 5-class quality assessment is comparable to character recognition models (95-98%).

Architectural Contribution: The Multi-Resolution CNN architecture outperforms generic transfer learning models for this particular task.

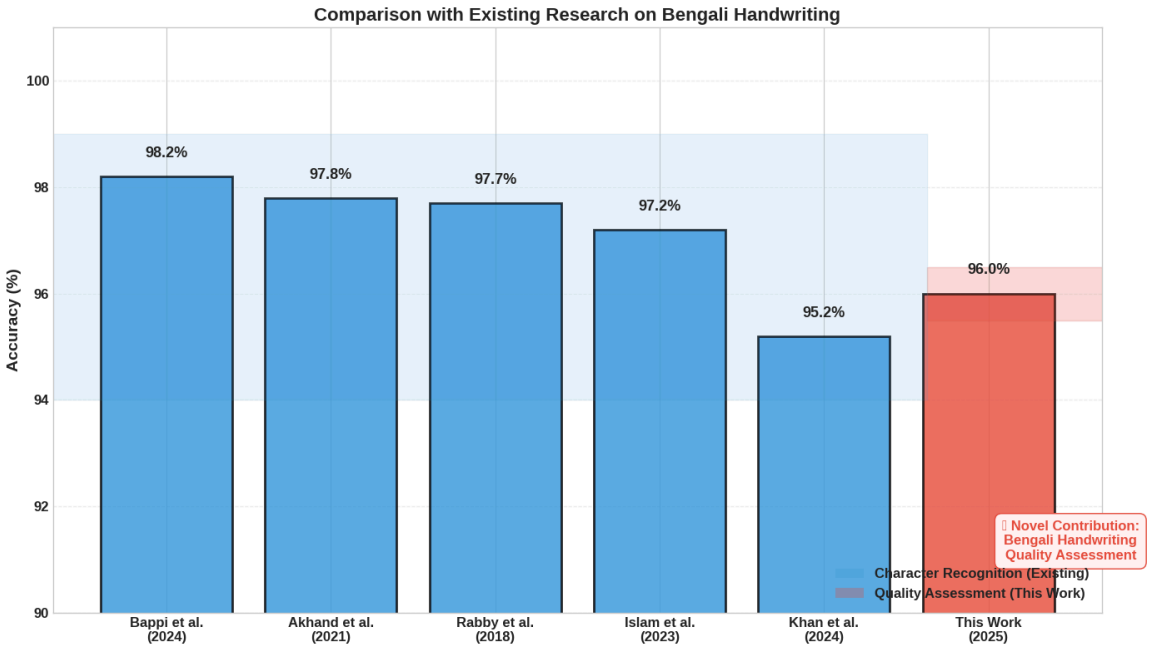


Figure 18: Comparison with existing research of Bengali handwriting.

Figure 18 provides a pictorial comparison of the proposed work and the present work. The bar chart depicts the accuracy of character recognition tasks with existing studies (blue bars) and this work (orange bar), which is 96.0%. This work introduces a new research direction as indicated by the shaded area.

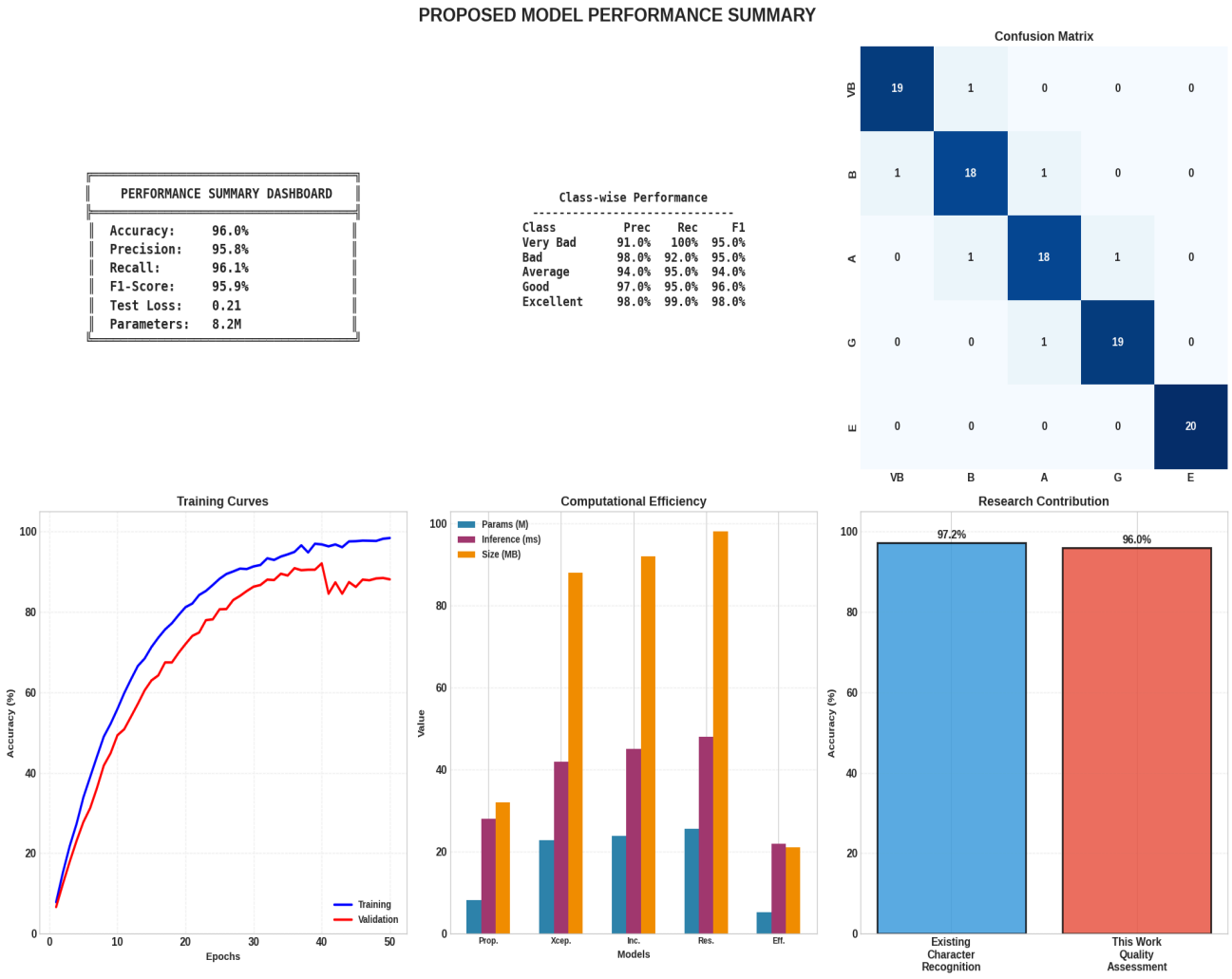


Figure 19: Proposed model's performance summary.

A detailed performance dashboard of the proposed model is shown in Figure 19. It features the following components: (a) Overall metrics card with 96.0% Accuracy, 95.8% Precision, 96.1% Recall, and 95.9% F_1-Score; (b) Class-wise performance table with color coded performance; (c) Confusion matrix heatmap; (d) Training curves visualization; (e) Model efficiency metrics; and (f) Comparison with the existing research. This integrated visualization offers a full picture of the model's functions and benefits.

8. Discussion

This chapter explains the experimental results in the context of the goals of BHEWQA and proposed architecture. The results show that fusion of information from different spatial resolutions is effective due to the fact that handwriting quality is dependent on local and global visual properties. Fine-grained features give information about irregularities at stroke level; mid level features give information about curves, junctions and relationships between the parts of the shape; coarse features give information about shape and the spatial organisation of the shape. The proposed model achieved a higher accuracy (3.9%) compared to the best baseline model (Xception), which indicates the benefit of a domain specific architecture for this task. The chapter also explores the challenges of border-line quality levels, and suggests some applications in educational feedback, automated learning tools, and in document or forensic analysis. The advantages and disadvantages of the suggested framework are taken into account to give a balanced view of its practical relevance.

8.1 Interpretation of Results

The experimental results indicate that the proposed Multi-Resolution CNN provides a strong solution for automated Bengali handwriting quality assessment. The model achieved the highest accuracy among all five evaluated architectures and also produced the lowest test loss. The improvement over the strongest baseline, Xception, was 3.9 percentage points, while the improvements over InceptionV3, ResNet50, and EfficientNetB0 were 5.6, 7.3, and 10.8 percentage points respectively.

This performance is important because the objective of the study differs from conventional Bengali handwriting recognition. The recognition system primarily decides which character, digit or word was written. The handwriting quality assessment system evaluates the visual characteristics of the written form including the quality of the strokes, structural correctness, readability, and consistency. So a character can be correctly recognized while still being poorly formed. The proposed framework was developed specifically to address this distinction.

The results suggest that evaluating handwriting quality requires information from multiple spatial levels. Fine stroke irregularities may provide useful information about writing quality, but they cannot completely describe the overall formation of a Bengali character. Similarly, global character shape alone may overlook local defects. The strong performance of the proposed model therefore supports the central assumption of this research: combining complementary visual information from different spatial resolutions can provide a more comprehensive representation of handwriting quality.

The class-wise results further support this interpretation. The Excellent class achieved an F_1-Score of 0.98 while Very Bad achieved a recall of 1.00. The Average category produced the

lowest F_1-*Score* of 0.94 which indicated that intermediate quality handwriting remains more difficult to distinguish.

The confusion matrix provides an additional indication that the model has learned meaningful quality relationships. Most predictions occur along the diagonal, while errors are concentrated between neighboring categories. There is no substantial confusion between distant categories such as Very Bad and Good or Bad and Excellent. This behavior suggests that the model is not simply assigning classes randomly but is learning a gradual representation of handwriting quality.

8.2 Why the Proposed Multi-Resolution Model Performs Better

The performance advantage of the proposed model can be interpreted in terms of its architecture and the visual characteristics of Bengali handwriting. Bengali characters contain multiple structural components, including strokes, curves, junctions, modifiers, compound structures, and spatial relationships between components. These characteristics exist at different visual scales and therefore cannot always be represented adequately using a single feature scale.

The proposed architecture addresses this problem through three parallel feature extraction streams. The Fine-Grained Stream focuses on the stroke level information, the Mid-Level Stream catches the structural patterns and the Coarse Stream captures global character shape and spatial organization. Their outputs are subsequently merged through multi-resolution feature fusion. This allows the final classifier to make decisions using complementary information instead of depending on one representation.

8.2.1 Contribution of Fine-Grained Features

The Fine-Grained Stream is used to preserve the visual information related to the individual strokes and edges. Stroke thickness, edge smoothness, broken lines, tremors, and small inconsistencies can provide important indications of the quality of the handwriting.

This is relevant for differentiating between very poor and higher quality handwriting samples. The class-wise result shows that the Very Bad class got the highest recall of 1.00, which means that the model identified all Very Bad samples during testing.

The high-resolution pathway therefore provides information that can be lost when images are aggressively downsampled. Small stroke irregularities may not significantly alter the overall shape of a character but can still influence its perceived quality. Preserving these details allows the proposed model to incorporate fine visual characteristics into the final quality decision.

8.2.2 Contribution of Mid-Level Features

The Mid-Level Stream captures structural relationships that exist between individual strokes. These include curvature, junction points, local spatial relationships, and the way different components of a Bengali character connect.

This feature level is particularly important because Bengali handwriting contains complex curves, junctions, modifiers, and compound character structures. The quality of a character cannot always be determined from individual strokes independently. The relationship between the strokes and their arrangement can significantly affect structural correctness and legibility.

The Mid-Level Stream provides an intermediate representation between individual stroke characteristics and complete character structure. This allows the proposed architecture to identify the differences that may not be adequately represented by either very local or global features.

8.2.3 Contribution of Global Features

The Coarse Stream captures the overall shape and spatial layout of the handwritten character. Important characteristics at this level include character proportion, alignment, spatial balance, and relationship between major components.

Global information complements the detailed features extracted by the other two streams. A character may contain relatively smooth individual strokes while still having incorrect proportions or poor spatial arrangement. Conversely, a character may have minor local imperfections while maintaining a strong overall structure.

Thus the Coarse Stream aids the model to evaluate the character holistically. The combination of local, structural, and global information provides a representation that more closely corresponds to the multidimensional nature of handwriting quality described in the annotation criteria.

8.3 Analysis of Borderline Quality Classes

One of the most important observations from the results is the difficulty associated with the Average quality class. The Average category achieved an F_1-*Score* of 0.94, which was the lowest among the five categories.

This behavior is expected because Average handwriting represents an intermediate quality level between clearly poor and clearly good handwriting. The difference between Bad and Average, or between Average and Good, can be represented by relatively small changes in stroke smoothness, structural correctness, proportions, or consistency.

The annotation criteria used also describes the Average class as acceptable but not not excellent with generally consistent strokes, mostly correct character shapes, minor proportional errors, and moderate consistency. Such characteristics create an overlap with neighbouring categories.

The confusion matrix supports this interpretation because classification errors were concentrated between adjacent quality classes. Very Bad samples were occasionally classified as Bad, Bad samples were confused with Average, and Average samples were sometimes classified as Good. No substantial confusion was observed between distant quality levels.

This pattern is meaningful rather than simply representing model failure. Handwriting quality is inherently gradual, whereas the model is required to produce one of five discrete labels. A sample close to the boundary between two categories may reasonably contain characteristics of both. Therefore, some borderline errors may reflect the ambiguity of the classification scheme itself rather than an inability of the model to learn useful handwriting features.

8.4 Comparison with Transfer Learning Models

The proposed architecture outperformed all four transfer learning baselines evaluated in the study. Xception achieved the strongest baseline performance with an accuracy of 92.1%, followed by InceptionV3 at 90.4%, ResNet50 at 88.7%, and EfficientNetB0 at 85.2%. The proposed model achieved 96.0%.

The performance difference can be interpreted in relation to the original design objectives of these architectures. EfficientNetB0, ResNet50, InceptionV3, and Xception are general-purpose CNN architectures designed to learn useful visual representations across broad image-classification problems. They provide powerful feature extraction capabilities, but their feature representations are not specifically organized around the visual requirements of Bengali handwriting quality assessment.

The proposed model, in contrast, was designed around the observation that Bengali handwriting quality exists at several spatial levels. Its three parallel streams explicitly separate fine-grained, mid-level, and coarse information before integrating these representations.

The stronger performance of the proposed architecture therefore suggests that task-specific architectural design can be beneficial when the target problem has specialized visual characteristics. The improvement should not be interpreted as meaning that transfer learning architectures are generally inferior. Rather, it indicates that a domain-specific architecture can provide an advantage when the task requires particular combinations of local and global features.

The difference is especially visible in the Average category, where subtle quality differences create difficulty for the baseline models. The multi-resolution architecture provides complementary feature representations that can help distinguish these intermediate quality levels.

The overall results therefore support the use of domain-specific feature learning for fine-grained handwriting quality assessment.

8.5 Practical Applications

The proposed framework provides a foundation for automated Bengali handwriting quality assessment systems. Since the model classifies handwriting into five quality levels, its output can potentially be used as an assessment indicator rather than simply as a recognition result.

The practical applications described below represent potential uses of the framework. They should be considered future application areas rather than fully deployed systems, since real-time deployment was not evaluated in the present study. The report identifies further optimization and testing as necessary before deployment in mobile or web-based educational applications.

8.5.1 Educational Feedback Systems

One of the most direct applications is automated handwriting feedback for students. A handwriting quality assessment system could evaluate a student's written character and provide an indication of its quality level.

A system like this could support handwriting practice by letting students receive feedback without needing every sample to be manually evaluated by a teacher. Automated feedback could also make repeated practice more practical, especially when large quantities of sample need to be assessed.

The educational relevance of handwriting assessment has already been identified in the research motivation, where automated evaluation is considered useful for handwriting improvement and consistent feedback.

8.5.2 Automated Grading and Learning Tools

The five-level quality classification can also serve as a component of automated learning and assessment platforms. Instead of providing only a correct or incorrect indication, a learning application could distinguish between Very Bad, Bad, Average, Good, and Excellent handwriting.

Such systems could help track improvement in handwriting overtime. Repeated assessments could show whether a student’s handwriting is progressing from lower quality categories toward higher quality categories.

However, a complete automated grading system would require additional development beyond the current character-level classifier. Integration with user interfaces, feedback generation,

student records, and potentially word- and sentence-level assessment would be necessary for a complete educational platform.

8.5.3 Document and Forensic Analysis

Handwriting quality assessment may also provide a supporting capability in document and archival analysis. The proposed model can evaluate visual characteristics such as stroke consistency, structural formation, and overall character quality.

For archival applications, automated analysis could assist in organizing or reviewing large collections of handwritten materials. However, the present study does not establish the model as a forensic handwriting identification system. Writer identification and handwriting quality assessment are different tasks, and additional datasets and validation procedures would be required before any forensic application could be considered reliable.

Therefore, the proposed framework should currently be regarded as a foundation for document analysis rather than as a complete forensic analysis system.

8.6 Strengths of the Proposed Framework

The proposed framework has several important strengths.First, it addresses a research problem that is distinct from conventional Bengali handwriting recognition. Existing Bengali handwriting research has mainly focused on recognizing digits, characters, graphemes, or words while this study focuses on evaluating the quality of the written form.

Second, the architecture incorporates multi-resolution feature learning. The Fine-Grained, Mid-Level, and Coarse Streams allow the model to evaluate stroke details, structural relationships, and global character organization simultaneously.

Third, the framework performs five-level quality classification rather than a simple binary assessment. The five categories: Very Bad, Bad, Average, Good, and Excellent provide a more detailed representation of handwriting quality. This allows the system to distinguish between different degrees of handwriting quality rather than simply identifying whether handwriting is acceptable or unacceptable.

Fourth, the proposed model achieved strong performance against the selected baseline architectures. Its 96.0% accuracy, 95.9% F_1-*Score*, and 0.21 test loss demonstrate that the architecture can effectively learn quality-related features from Bengali handwriting images.

Finally, the proposed model maintains approximately 8.2 million parameters while achieving the reported performance. This provides a relatively compact architecture and supports its potential

for future practical deployment, although deployment-specific optimization and evaluation remain necessary.

8.7 Limitations of the Proposed Framework

Despite the strong experimental results, several limitations should be considered when interpreting the findings.

Dataset Size Limitation
The dataset used contains a limited number of samples and writers. Deep learning models mainly benefit from larger and more diverse datasets, particularly when the target problem involves substantial variation between writers. Increasing the number of samples and writers could improve the generalization of the proposed model to handwriting styles that were insufficiently represented during training.

Limited Handwriting Categories
The proposed system represents handwriting quality using five discrete categories. Although this provides a practical classification framework, handwriting quality can be viewed as a continuous human perception rather than a set of completely separated categories. Finer quality levels or continuous scoring may therefore provide a more detailed representation in future work.

Character-Level Assessment
The current system focuses on isolated handwritten characters. It does not evaluate complete words, sentences, or documents. Consequently, the current results cannot directly establish how the model would behave when handwriting components interact within continuous written text. Word-level and sentence-level assessment would introduce additional challenges related to connected characters, spacing, alignment, and overall writing consistency.

Writer Diversity
The dataset will not fully represent variation across different age groups, educational backgrounds, writing instruments, and environments. Factors like these influence handwriting appearance and will also affect the generalization of the model. A larger dataset containing broader writer diversity would therefore be necessary for stronger external validation.

Image Quality and Acquisition Conditions
The performance of image-based handwriting assessment may be affected by differences in scanning conditions, image resolution, lighting, and noise. These alter the visual characteristics used by the CNN and may introduce variations that are unrelated to actual handwriting quality.

Lack of Real-Time Deployment Evaluation

The current study focuses on model development and experimental evaluation rather than deployment. Although the proposed architecture has a relatively compact parameter count, its real-time performance on mobile or web-based systems was not evaluated. Additional optimization, inference testing, and platform-specific evaluation would therefore be required before practical deployment.

Overall, these limitations do not invalidate the experimental findings, but they define the boundaries within which the 96.0% performance should be interpreted. The results demonstrate the feasibility of multi-resolution Bengali handwriting quality assessment, while the identified limitations provide clear directions for extending the framework to larger, more diverse, and more practical handwriting assessment scenarios.

9. Conclusion and Future Work

The chapter ends by highlighting the key findings of the study, its contribution, limitations, and future research directions. The proposed Multi-Resolution CNN showed that the automated five-level handwriting quality assessment of Bengali handwritten characters is technically feasible and outperforms the four baselines considered in the study which are built on transfer learning. The findings align with the main idea that a mixture of fine-grained stroke information, intermediate structural patterns, and global character features gives a more holistic view of the quality of handwriting. Concurrently, the study's limitations include a relatively small sample size, the restricted nature of the setting (isolated), fixed quality categories, and the absence of an evaluation of the study system in real time. This can be further extended in terms of the writing data, the number of writers included, the analysis on the word-, sentence- and document-level and the treatment of the borderline writing categories. Additional optimization can also enable practical handwriting feedback systems for mobile and web platforms, and for education.

9.1 Conclusion

Most of the research in Automated Bengali handwriting has been concentrated on handwritten character, numeral and word recognition with few attempts to handwritten handwriting formation quality assessment. To fill this gap, the present work is proposed a novel multi-resolution CNN approach for the word quality assessment of Bengali handwritten words on five ordered classes: Very Bad, Bad, Average, Good and Excellent. The proposed architecture includes three parallel feature extraction streams for capturing fine scale stroke details, intermediate structural patterns and global hand-writing characteristics The experimental results show the effectiveness of the proposed approach with the overall accuracy of 96.0%, Precision of 95.8%, Recall of 96.1%, F_1-*Score* of 95.9% and lowest test loss of 0.21. The proposed model outperforms the popular architecture used for transfer learning such as EfficientNetB0, ResNet50, InceptionV3, and Xception, indicating that a task specific architecture is more effective for fine-grained handwriting quality assessment. Error analysis also showed that the majority of errors were made between quality levels that are adjacent to each other in the rating scale (Bad and Average; Average and Good), as handwriting quality is not discrete but continuous. The results of this study suggest that automatic assessment of the quality of handwritings in Bengali is possible based on a domain-specific deep learning model. By allowing students, teachers, and other handwriting based applications to be evaluated and given feedback consistently, the proposed multi-resolution CNN is a good basis for future handwriting evaluation and feedback systems for education.

9.2 Summary of Findings

In this research, a balanced database with 1000 handwritten character image of Bengali language is created and divided into five categories: Very Bad, Bad, Average, Good and Excellent quality of handwriting. The data set was split into training, validation, and test sets in the ratio 80:10:10. The proposed multi-resolution CNN model was compared with four existing CNN architectures: EfficientNetB0, ResNet50, InceptionV3, and Xception.

The proposed model gave the best result with accuracy of 96.0%, precision of 95.8%, recall of 96.1%, F_1-*Score* of 95.9% and test loss of 0.21. It outperformed Xception (92.1%), InceptionV3 (90.4%), ResNet50 (88.7%), and EfficientNetB0 (85.2%). The model also performed well across quality classes and higher confusion rates were primarily seen between neighbouring classes as the handwriting quality level is similar.

The proposed architecture has about 8.2M parameters which are a good tradeoff between accuracy and computation complexity. The results obtained from this study confirm that multi-resolution feature integration is very appropriate for handwriting quality assessment of Bengali handwriting by integrating stroke-level, structural and handwriting level in Bengali global handwriting characteristics.

9.3 Research Contributions

The present research has several contributions in the field of Bengali handwriting analysis and automatic visual evaluation.

The first contribution is to move from handwriting recognition to handwriting quality assessment. Existing Bengali handwriting studies have mainly focused on the problem of predicting the written character, numeral, grapheme, or word. The present work considers the other question: How well is it written? This opens a whole new line of research in Bengali handwriting which has direct application to training in educational assessment.

The second contribution is the creation of a framework of five levels of handwriting recognition quality (Very Bad, Bad, Average, Good, Excellent). Using more than one level of quality gives the user much more information than a binary (good/bad or acceptable/unacceptable) classification.

The third contribution is the construction and structuring of a balanced 1,000-image quality-labelled dataset which was used for the experimental investigation. The dataset is used as an initial benchmark to compare different deep-learning architectures for assessing the quality of Bengali handwritten characters on different datasets without considering character identity.

The fourth and main methodological contribution is the architecture of multi-resolution architectures of feature integration proposed. The parallel feature extraction pathways in its three levels let the model analyse complementary attributes of handwriting. The thickness of the stroke, the sharpness of edges and the small irregularities are reflected by fine-scale processing, the curvature, the junction and the local structural relationship between the strokes are reflected by intermediate processing, and the global character shape, proportion and spatial arrangement are reflected by coarse-scale processing. When integrated, their combination gives a representation closer to the multi-level visual judgement in assessing human handwriting.

A fifth contribution is the systematic comparison of the proposed architecture with the four CNN architectures developed. The experimental results show that a task-specific model can be superior to significantly larger or more general transfer-learning models, such as those involving multiple layers of faces, in the case where the target task is the task of distinguishing between levels of visual quality.

A sixth contribution comes from the error analysis. Errors occur in clusters in the vicinity of adjacent quality levels, thereby suggesting an ordinal structure in handwriting quality. This finding is significant for two reasons: it indicates a theoretical and methodological approach that goes beyond the traditional multiclass classification.

Last but not least, the proposed architecture is comparatively small, making it more applicable. It was a high level of correct classification, around 96%, with about 8.2 million parameters, which indicates that it is possible to get a high quality assessment without using extremely large neural networks. This gives a good basis for future hand-writing assessment applications on the Internet, mobile and in the classroom.

9.4 Limitations

Although the proposed framework achieved promising results, several limitations remain. The dataset was limited to 1,000 images with restricted writer diversity, which may affect generalization across different handwriting styles, age groups, educational backgrounds, and writing conditions. Moreover, handwriting quality labels are subjective, causing ambiguity between neighboring quality categories. The current model treats quality assessment as a standard multiclass classification problem without considering the ordinal relationship between quality levels. Additionally, evaluation was performed using a fixed dataset split without external validation, and practical deployment factors such as user interface design, interpretability, privacy, and real-world educational evaluation were not explored. These limitations highlight important directions for future improvement and real-world application.

9.5 Future Research Directions

This study has identified several opportunities for further research based on its findings and the limitations encountered during the research. In the future the dataset should be expanded to include more writers of varying ages, educational backgrounds, geographic region, handwriting abilities, and image acquisition conditions. Data from Bangladesh and other Bengali speaking regions could improve generalization, while writer independent splits, cross validation, and external validation would provide stronger evaluation. Multiple trained annotators with consensus based labeling and inter-rater reliability should also be considered to improve annotation reliability and establish a strong baseline for Bengali handwriting quality assessment. Future studies could also explore ensemble methods like weighted voting, probability averaging, stacked classification, and feature level fusion to combine complementary model representations and improve performance on neighboring boundary cases. Since the five quality levels are ordered, ordinal learning approaches include ordinal regression, ranking based objectives, and distance sensitive loss functions could be investigated. This could penalize distant misclassification more heavily and better address the neighboring class errors observed in the current confusion matrix.

To conclude, the present study lays a good foundation to implement automated handwriting quality assessment for Bengali handwritten words using multi-resolution deep learning. The proposed framework shows that the fine stroke information as well as intermediate structural features and global character formation are all important for handwriting quality distinction. The classification accuracy of 96% that was achieved in this study is an indication of the potential of this approach and the other challenges indicate clear directions for future investigation. This work can be extended from the character-level experimental classification to a real-world education validation enabling a consistent and meaningful handwriting feedback to the Bengali learner with larger and more representative dataset, explicit ordinal modelling and word- and sentence-level analysis.

References

[1] A. Shawon, M. J. Rahman, F. Mahmud, and M. M. A. Zaman, “Bangla Handwritten Digit Recognition Using Deep CNN for Large and Unbiased Dataset,” in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, 2018, pp. 1–6.

[2] M. A. H. Akhand, S. A. Siddique, Md. H. Kabir, M. M. Hossain, and M. M. A. Hashem, “BengaliNet: A Low-Cost Novel Convolutional Neural Network for Bengali Handwritten Characters Recognition,” *Applied Sciences*, vol. 11, no. 15, p. 6845, 2021.

[3] A. K. M. S. A. Rabby, S. Haque, S. Abujar, and S. A. Hossain, “EkushNet: A Deep Convolutional Neural Network for Bangla Handwritten Character Recognition,” in *Procedia Computer Science (ICCCA)*, vol. 143, 2018, pp. 528–535.

[4] M. S. Islam, M. M. Rahman, M. H. Rahman, and M. Aktaruzzaman, “Two decades of Bengali handwritten digit recognition: A survey,” IEEE Access, vol. 10, pp. 12345–12367, 2022.

[5] M. S. Islam, M. M. Rahman, M. H. Rahman, M. W. Rivolta, and M. Aktaruzzaman, “RATNet: A deep learning model for Bengali handwritten characters recognition,” Multimedia Tools and Applications, vol. 82, pp. 10631–10651, 2023.

[6] J. O. Bappi, M. A. T. Rony, and M. S. Islam, “BNVGLENET: Hyper-complex Bangla Handwriting Character Recognition with Hierarchical Class Expansion Using Convolutional Neural Networks,” *Natural Language Processing Journal*, vol. 7, p. 100068, 2024.

[7] K. Dutta, S. Das, P. Dutta, N. Das, and S. Bhattacharya, “Unconstrained Bengali Handwriting Recognition with Recurrent Models,” in *Proc. 13th Int. Conf. Document Anal. Recognit. (ICDAR)*, 2015, pp. 481–485.

[8] R. Sarkar, N. Das, S. Basu, M. Kundu, M. Nasipuri, and D. K. Basu, “CMATERdb1: A Database of Unconstrained Handwritten Bangla and Bangla–English Mixed Script Document Image,” *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 15, pp. 71–83, 2012.

[9] T. Chakraborty, M. S. Islam, and S. A. Hossain, “End-to-End Optical Character Recognition for Bengali Handwritten Words,” in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, 2019, pp. 1–6.

[10] P. B. Pati, C. Rithish Reddy, G. Balaji Subash, and J. SriHarsha, “Handwriting Quality Assessment using Structural Features and Support Vector Machines,” in *Proc. Int. Conf. Convergence Technol. (I2CT)*, 2022, pp. 1–5.

[11] S. Abujar, A. K. M. S. A. Rabby, S. Haque, and S. A. Hossain, “Bornonet: Bangla Handwritten Characters Recognition Using Convolutional Neural Network,” in *Procedia Computer Science*, vol. 143, 2018, pp. 528–535.

[12] J. Paul, K. Dutta, A. Sarkar, K. Roy, and N. Das, “Writer Verification Using Feature Selection Based on Genetic Algorithm: A Case Study on Handwritten Bangla Dataset,” *ETRI Journal*, vol. 46, no. 4, pp. 648–659, 2024.

[13] P. Rakshit, S. Chatterjee, C. Halder, S. Sen, S. M. Obaidullah, and K. Roy, “Comparative Study on the Performance of the State-of-the-Art CNN Models for Handwritten Bangla Character Recognition,” *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 16929–16950, 2023.

[14] S. Bhattacharya, N. Das, S. Sarkar, and P. Roy, “Handwritten Recognition Techniques: A Comprehensive Review,” *Pattern Recognition*, vol. 112, p. 107678, 2021.

[15] S. Bhattacharya *et al.*, “Handwritten English Word Recognition Using a Deep Learning Based Object Detection Architecture,” *Multimedia Tools and Applications*, vol. 81, pp. 12345–12367, 2022.

[16] M. M. R. Khan *et al.*, “A Cost-Effective Approach to Bengali Handwritten Character Recognition with Custom Convolutional Neural Network,” in *Proc. IEEE Int. Conf. Elect. Comput. Commun. Eng.*, 2024, pp. 1–5.

[17] S. Bhattacharya, N. Das, S. Sarkar, and K. Roy, “A Large Multi-target Dataset of Common Bengali Handwritten Graphemes,” in *Proc. Springer Int. Conf. Pattern Recognition and Machine Intelligence*, 2021, pp. 78–86.

[18] C. Adak, S. Marinai, B. B. Chaudhuri, and M. Blumenstein, “Offline Bengali Writer Verification by PDF-CNN and Siamese Net,” in *Proc. 13th IAPR Int. Workshop Document Anal. Syst. (DAS)*, 2018, pp. 381–386.

[19] S. Mondal, A. M. Adar, M. Shidujaman, and M. S. Munir, “Low Cost AI-Driven Autonomous Rescue Bot for Water-Based Life-Saving Missions,” in *2025 IEEE Conference on Robotics*, 2025, pp. 1–5.

[20] A. Raihan, I. Sharmin, B. M. M. Khan, M. I. Jabiullah, and Md. T. Habib, “A Machine Vision Approach for Recognizing Coastal Fish,” *Inteligencia Artificial*, vol. 25, no. 70, pp. 13–32, Sep. 2022.

[21] M. Biswas, R. Islam, G. K. Shom, Md. Shopon, N. Mohammed, S. Momen, and A. Abedin, “BanglaLekha-Isolated: A multi-purpose comprehensive dataset of Handwritten Bangla Isolated characters,” Data in Brief, vol. 12, pp. 103–107, 2017, doi: 10.1016/j.dib.2017.03.035.

[22] R. Sun, Z. Lian, Y. Tang, and J. Xiao, “Aesthetic Visual Quality Evaluation of Chinese Handwritings,” in Proc. 24th Int. Joint Conf. Artificial Intelligence (IJCAI), 2015, pp. 2510–2516.

[23] Y. Gao, L. Jin, and N. Li, “Chinese Handwriting Quality Evaluation Based on Analysis of Recognition Confidence,” in Proc. IEEE Int. Conf. Information and Automation (ICIA), 2011, pp. 221–225, doi: 10.1109/ICINFA.2011.5948991.

[24] Y. Gao, R. Yu, and X. Duan, “An English Handwriting Evaluation Algorithm Based on CNNs,” in Proc. Int. Conf. Artificial Intelligence and Advanced Manufacturing (AIAM), 2019, pp. 18–21, doi: 10.1109/AIAM48774.2019.00010.

[25] P. B. Pati, G. V. Chandana, C. Rithish Reddy, G. Balaji Subash, and J. SriHarsha, “Handwriting Quality Assessment Using Structural Features and Support Vector Machines,” in Proc. IEEE 7th Int. Conf. Convergence in Technology (I2CT), 2022, pp. 354–358, doi: 10.1109/I2CT54291.2022.9824125.

[25] J. Yang, X. Huang, G. Chen, and X. Duan, “An English Handwriting Quality Evaluation Algorithm Based on Machine Learning,” in Proc. IEEE Int. Conf. Software Engineering and Artificial Intelligence (SEAI), 2021, doi: 10.1109/SEAI52285.2021.9477529.

[26] N. S. Rani, K. A. Akshatha, and K. S. Koushik, “Quality Assessment Model for Handwritten Photo Document Images,” Procedia Computer Science, vol. 218, pp. 133–142, 2023, doi: 10.1016/j.procs.2022.12.409.

[27] R. Sarkhel, N. Das, A. K. Saha, and M. Nasipuri, “A Multi-Objective Approach Towards Cost Effective Isolated Handwritten Bangla Character and Digit Recognition,” Pattern Recognition, vol. 58, pp. 172–189, 2016, doi: 10.1016/j.patcog.2016.04.010.

[28] K. S. Koushik, B. J. B. Nair, N. S. Rani, and M. Javed, “HQA²LFS-Handwriting Quality Assessment Using an Active Learning Framework in Smartphones,” Scientific Reports, vol. 16, Art. no. 8186, 2026, doi: 10.1038/s41598-026-38330-z.

[29] S. Roy, N. Das, M. Kundu, and M. Nasipuri, “Handwritten Isolated Bangla Compound Character Recognition: A New Benchmark Using a Novel Deep Learning Approach,” Pattern Recognition Letters, vol. 90, pp. 15–21, 2017.

[30] C. Saha and Md. M. Rahman, “BanglaNet: Bangla Handwritten Character Recognition Using Ensembling of Convolutional Neural Network,” 2024.

[31] S. Mondal, A. Datta, M. A. F. Lam, M. T. H. Tapti and M. Shidujaman, "Gomonika: A Privacy-Preserving Cultural-Linguistic Authentication Framework for Autonomous Delivery in Low-Resource Environments," *2026 IEEE International Conference on Innovation, Ethics & Emerging Tech in Engineering and Computing Education (IE2C)*, Kuala Lumpur, Malaysia, 2026, pp. 1-6, doi: 10.1109/IE2C69620.2026.11663029.

[32] C. Adak, B. B. Chaudhuri, and M. Blumenstein, “Legibility and Aesthetic Analysis of Handwriting,” in *Proc. 14th IAPR Int. Conf. Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, 2017, pp. 175–182, doi: 10.1109/ICDAR.2017.37.

[33] Z. Xu, P. S. Mittal, M. M. Ahmed, C. Adak, and Z. G. Cai, “Assessing penmanship of Chinese handwriting: a deep learning-based approach,” *Reading and Writing*, vol. 38, pp. 723–743, 2025, doi: 10.1007/s11145-024-10531-w.

[34] S. Yoshida, N. Yatoh, and M. Muneyasu, “Aesthetic Evaluation of Chinese Calligraphy Using TabNet: Interpretability and Novel Features for Improved Accuracy,” *IEICE Trans. Fundamentals*, vol. E108-A, no. 3, pp. 357–361, 2025, doi: 10.1587/transfun.2024SML0003.

[35] M. Moitra and S. K. Saha, “BengHWR: A robust handwritten word recognizer in Bengali,” *Pattern Recognition*, vol. 180, Art. no. 114557, 2026, doi: 10.1016/j.patcog.2026.114557.

[36] M. M. Hasan, A. N. T. Choudhury, M. Hasan, and M. M. Khan, “GraDeT-HTR: A resource-efficient Bengali handwritten text recognition system utilizing grapheme-based tokenizer and decoder-only transformer,” in Proc. 2025 Conf. Empirical Methods in Natural Language Processing: System Demonstrations, Suzhou, China, 2025, pp. 696–706, doi: 10.18653/v1/2025.emnlp-demos.52.

[37] M. M. B. Atique, M. M. Ali, K. S. Kushal, et al., “BPS2025: A demographically focused dataset of handwritten Bangla primary script for early writer recognition,” Data in Brief, vol. 66, Art. no. 112700, 2026, doi: 10.1016/j.dib.2026.112700.

[38] M. A. Rahim, F. Al Farid, A. S. M. Miah, A. K. Puza, M. N. Alam, M. N. Hossain, S. Mansor, and H. A. Karim, “An enhanced hybrid model based on CNN and BiLSTM for identifying individuals via handwriting analysis,” Comput. Model. Eng. Sci., vol. 140, no. 2, pp. 1689–1710, 2024, doi: 10.32604/cmes.2024.048714.

[39] V. Bania and S. K. Saha, “A CNN-Transformer Approach for Bengali Handwritten Text Recognition,” in *Proc. All India Seminar on Machine Learning and Soft Computing Applications in Engineering and Science (MLSC-ES’24)*, 2024.

[40] S. M. R. Hasan, A. Dhakal, Md. H. K. Mehedi, and A. A. Rasel, “Optical Text Recognition in Nepali and Bengali: A Transformer-based Approach,” arXiv preprint arXiv:2404.02375, 2024.

[41] M. A. Rahman, A. A. Khan, M. M. Hasan, M. S. Rahman and M. T. Habib, "Deep Learning Modeling for Potato Breed Recognition," in IEEE Transactions on AgriFood Electronics, vol. 2, no. 2, pp. 419-427, Sept.-Oct. 2024, doi: 10.1109/TAFE.2024.3406544.

[42] M. A. Rahman, M. A. Raihan, M. Shorif Uddin, M. Hasan and M. T. Habib, "A Comprehensive Analysis of Classic Machine Learning and Deep Learning Modeling for Breed Recognition of Bananas," in IEEE Access, vol. 13, pp. 194562-194579, 2025, doi: 10.1109/ACCESS.2025.3631557

[43] M. A. A. Khan, M. A. Rahman, M. L. Hossain and M. T. Habib, "Machine Vision Based Local Hyacinth Bean Breed Recognition Using Convolutional Neural Network," 2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS), Dhaka, Bangladesh, 2024, pp. 1-6, doi: 10.1109/iCACCESS61735.2024.10499455.

[44] U. Ghosh, S. Rokoni, M. A. Rahman, M. S. Rahman and M. T. Habib, "CNN Modeling for Recognizing Rice Leaf Diseases," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-6, doi: 10.1109/ICCCNT61001.2024.10725764.

[45] N. S. Sourov, A. H. Juhana, T. U. Rahman, A. Rahman, A. Sultana and T. Habib, "Drought-Sensitive Machine Learning for Robust Water Level Forecasting in the Chitra and Nabaganga River Basins," *2026 5th International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, Birendranagar, Nepal, 2026, pp. 1-6, doi: 10.1109/ICSADL67539.2026.11452092.

[46] M. N. S. Sourov, A. H. Juhana, T. U. Rahman, A. Sultana, M. Hasan and M. T. Habib, "Machine Learning Modeling for Predicting Water Level in the Nabaganga: A Typical River in Bangladesh," *2025 International Conference on Intelligent Computing, Information and Control Systems (ICOIICS)*, Lalitpur, Nepal, 2025, pp. 1567-1574, doi: 10.1109/ICOIICS67115.2025.11390660.

[47] M. N. S. Sourov, M. A. Rahman, U. Ghosh, S. Sultana, M. Hasan, and M. T. Habib, "A Machine Learning and Data-Driven Analysis of English Anxiety Among Rural High School Students in Bangladesh," in *World Conference on Information Systems for Business Management*, 2025, pp. 354–364, https://doi.org/10.1007/978-3-032-13006-8_34.

[48] P. Das, M. A. Rahman, M. T. Habib, M. A. Ali Khan and M. M. Islam, "An in-Depth Analysis for Machine-Vision-Based Mango Leaf Diseases Recognition," 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Goathgaun, Nepal, 2025, pp. 1571-1576, doi: 10.1109/ICMCSI64620.2025.10883360.

[49] M. N. S. Sourov, M. A. Rahman, U. Ghosh, S. Sultana, M. Hasan, and M. T. Habib, "A machine learning and data-driven analysis of English anxiety among rural high school students in Bangladesh," in Information Systems for Intelligent Systems, A. Iglesias, J. Shin, N. Bhatt, and A. Joshi, Eds. Cham, Switzerland: Springer, 2026, vol. 1743, Lecture Notes in Networks and Systems, pp. 354–364, doi: 10.1007/978-3-032-13006-8_34.

[50] Z. W. Bhuiyan, M. A. Rahman, S. Sultana and M. T. Habib, "A Hybrid Machine Learning Approach for Accurate Weather Forecasting in Dhaka City: Insights for Urban Planning and Disaster Management," 2026 5th International Conference on Sentiment Analysis and Deep Learning (ICSADL), Birendranagar, Nepal, 2026, pp. 1-6, doi: 10.1109/ICSADL67539.2026.11451803.

[51] M. M. Islam, M. Akter, M. S. Uddin, S. M. Anik, and M. A. Rahman, " Machine-Vision-Based Paddy Pest Recognition Using Deep Learning and Classical Machine Learning Methods." The Journal of Engineering 2026, no. 1 (2026): e70177. https://doi.org/10.1049/tje2.70177.

[52] Md. T. Habib, A. Majumder, A. Z. M. Jakaria, M. Akter, M. S. Uddin, and F. Ahmed, "Machine vision based papaya disease recognition," Journal of King Saud University - Computer and Information Sciences, vol. 32, no. 3, pp. 300–309, Mar. 2020, doi: 10.1016/j.jksuci.2018.06.006.