ALGERIAN DEMOCRATIC AND POPULAR REPUBLIC

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

KASDI MERBAH UNIVERSITY OUARGLA

FACULTY OF NEW INFORMATION AND COMMUNICATION TECHNOLOGIES

DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

THESES SUBMITTED IN CANDIDACY FOR A MASTER DEGREE IN COMPUTER SCIENCE, OPTION INDUSTRIAL

BY NOUSSAIBA KHELLAF & Zohra Hadjer Guemri

THEME

Automatic Recognition of Ancient Arabic Manuscript Authors Using Deep Learning

JURY MEMBERS

Mrs. Leila Amrane	Supervisor	UKM Ouargla
Dr. Zineb Kaoudja	Jury Chair	UKM Ouargla
Dr. Aicha Korichi	Examiner	UKM Ouargla

ACADEMIC YEAR: 2024/2025

Acknowledgment

First and foremost, we would like to express our deepest gratitude to the One above all, the omnipresent God, for answering our prayers and granting us the strength and perseverance to overcome challenges throughout this journey. Thank You, God.

We extend our sincere and profound appreciation to our supervisor, Dr. Laila Imrane, for her unwavering support, insightful guidance, constructive feedback, and continuous encouragement. Her expertise and commitment have been instrumental in shaping the direction of this research and in overcoming the obstacles we encountered along the way. We are especially grateful for the trust and confidence she placed in us.

We are also deeply thankful to the esteemed members of the examination committee for accepting to evaluate our work. It is a great honor to have you review our thesis, and we highly value your time, insights, and forthcoming comments.

Our heartfelt thanks go to all the faculty members and instructors who have contributed to our academic journey. Their dedication to teaching and their encouragement have played a vital role in our growth, both intellectually and personally.

Words cannot fully express our gratitude to our beloved parents. To my late father, whose memory remains a guiding light and a source of inspiration, and to my loving mother, whose unconditional love and support continue to sustain me—thank you. Your sacrifices, faith, and encouragement have been the foundation of our success.

We would also like to extend special thanks to our families and friends, whose steadfast support and belief in us have provided immense emotional strength throughout this endeavor.

Finally, we are grateful to every individual who, in one way or another, contributed to the completion of this thesis. Your encouragement, support, and kindness have helped us reach this significant milestone

etoolbox fancyhdr

Dedication

Praise be to Allah. Alhamdulillah for everything. I am honored to dedicate this graduation to my parents, whose unwavering support and values have guided me to this point. To my late father, whose memory inspires me, and my dear mother, who taught me to write without regret. May God bless her with a long and prosperous life. I extend my heartfelt dedication to my teachers and my Department for their education and support. Special thanks to my supervisor, Mrs. Leila Amrane, for his invaluable guidance and encouragement throughout this journey. Lastly, to my family and friends who have stood by me, thank you for your love and support. This work is dedicated to all of you. NOUSSAIBA KHELLAF

Dedication

Praise be to Allah, first and foremost, for every blessing—visible and hidden—and for every success and ease along the way.

I dedicate this humble work to those who instilled in me the values of determination and sincerity: my beloved parents. To the soul of my late father, whose kind memory continues to inspire and strengthen me, and to my dear mother, who taught me to write with confidence and move forward without regret—may Allah bless her with long life, good health, and abundant happiness.

I also dedicate this graduation to my esteemed teachers and the Department of Computer Science, in deep appreciation for the knowledge, guidance, and support they provided throughout my academic journey.

A special note of gratitude goes to my supervisor, Mrs. Leila Amrane, whose dedication and insightful guidance played a pivotal role in the completion of this work. I am truly grateful for his support.

To my family and loyal friends, who have always stood by my side, I offer my heartfelt thanks for your genuine love and unwavering encouragement.

This achievement is the fruit of our collective efforts, and I proudly dedicate it to all of you.

Zohra Hadjer Guemri

Abstract

This study investigates the effectiveness of deep learning techniques in the automatic identification of authors of ancient Arabic manuscripts. The motivation lies in preserving Arab cultural heritage and supporting academic research in the field of palaeography. Given the absence of ready-made, task-specific datasets, a custom dataset was manually compiled, comprising 10 authors, each represented by 20 manuscript images. This dataset served as a starting point to assess the feasibility of authorship classification based on handwriting patterns. To prepare the data for model training, a series of preprocessing steps and data augmentation techniques were applied to improve input quality and reduce overfitting risk. The study implemented a simple Convolutional Neural Network (CNN) model and evaluated three pre-trained architectures: VGG16, InceptionV3, and ResNet50. The results demonstrated that the custom CNN model achieved a test accuracy of 90%, while VGG16 and InceptionV3 reached 95% due to the advantages of transfer learning. In contrast, ResNet50 performed less effectively, achieving 75% accuracy, likely due to its complexity and higher data requirements. The study confirms that deep learning models can capture distinctive handwriting features even from relatively small datasets. These findings support the potential of AI-based systems in historical manuscript analysis and suggest future directions, such as expanding the dataset, integrating metadata, and developing practical tools for automated authorship attribution in Arabic script manuscripts.

Keywords: Arabic Manuscripts, Authorship Attribution, Deep Learning, CNN, Transfer Learning, Palaeography

Résumé

L'attribution d'auteur des manuscrits anciens constitue un problème complexe à l'intersection de l'analyse historique et des méthodes computationnelles. Les approches traditionnelles reposent largement sur des processus d'examen manuel dirigés par des experts, qui sont intrinsèquement chronophages et nécessitent une expertise spécifique au domaine. Avec l'avancée rapide de l'apprentissage automatique, en particulier des méthodes d'apprentissage profond, il existe un potentiel croissant pour automatiser l'analyse de l'écriture manuscrite historique. Ces approches permettent non seulement d'améliorer la précision de la classification, mais aussi de réduire considérablement les ressources humaines et computationnelles nécessaires pour les tâches d'identification d'auteur. La plupart des études précédentes utilisent des manuscrits étrangers rédigés en latin, avec une absence totale de la langue arabe. Cette étude vise à identifier les auteurs de manuscrits arabes anciens à l'aide de l'apprentissage profond.

Mots-clés : Analyse de documents, Manuscrits anciens, Apprentissage profond, Reconnaissance optique de caractères

Acronyms and Technical Terms

Acronyms

Acronym	**Full Form**
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
OCR	Optical Character Recognition
GT	Ground Truth
HWR	Handwriting Recognition
GPU	Graphics Processing Unit
SVM	Support Vector Machine
HOG	Histogram of Oriented Gradients
HMM	Hidden Markov Model
IAM	IAM Handwriting Database
CVL	CVL Handwriting Database
API	Application Programming Interface
NLP	Natural Language Processing

Technical Terms

Term	**Explanation**
Handwriting	The act or style of writing by hand.
Author Identification	Determining the writer of a given manuscript or text.
Writer Retrieval	Finding documents written by the same writer.
Palaeography	The study of ancient handwriting and scripts.
Feature Extraction	Identifying relevant characteristics from data.
Preprocessing	Initial steps to clean or prepare data for analysis.
Segmentation	Dividing an image or text into meaningful parts (e.g., lines, words, characters).
Script	A specific style or system of writing (e.g., Arabic, Latin).
Calligraphy	Artistic, stylized, or elegant handwriting or lettering.
Dataset	A structured collection of data used for training or evaluation.

Contents

List of Figures

List of Tables

General Introduction

I. Research Problem:

The process of identifying the authors of ancient manuscripts represents a real challenge due to the deterioration these documents have undergone and the diversity of handwriting styles and methods from one author to another. Traditional approaches to manuscript analysis rely heavily on human expertise, which can often be slow and inaccurate. With the emergence of artificial intelligence technologies, particularly deep learning, new possibilities have arisen for developing tools capable of automatically identifying the writing characteristics of different authors. However, this approach raises several questions: How can deep learning techniques, especially Convolutional Neural Networks (CNNs), be utilized to build an effective and accurate system capable of automatically identifying the authors of ancient Arabic manuscripts? Furthermore, what methodologies should be adopted to address challenges such as image degradation and handwriting variability to ensure the accuracy of the results?

II. Research Objectives :

This research aims to develop a comprehensive system capable of classifying and analyzing Arabic manuscripts to identify their authors using deep learning techniques. This objective represents a practical contribution to the field of studying ancient Arabic manuscripts, as it will help improve the accuracy of identifying the authors of manuscripts through advanced techniques such as Convolutional Neural Networks (CNN). In other words, the research seeks to study the deep learning classification technique (CNN), which can assist researchers in identifying the authors of manuscripts in general. However, it specifically aims to work on a CNN model to analyze Arabic manuscripts and accurately determine their authors.

Goals to be Achieved

- Develop an automated system capable of attributing authorship to ancient Arabic manuscripts using deep learning techniques.
- Bridge the gap in current research by focusing on Arabic-language manuscripts, which are largely absent in existing studies.
- Improve classification accuracy in authorship attribution through the application of modern neural network architectures.
- Reduce reliance on manual expert analysis by introducing scalable, data-driven solutions.
- Build a benchmark dataset of digitized Arabic manuscripts to support future research in the field.
- Contribute to the preservation and understanding of Arabic literary heritage through computational means.

III. Organization of the project :

Chapter 1: Identification of Ancient Manuscript Authors
This chapter introduces the project on the automatic identification of ancient manuscript authors using images. It discusses the historical context, the significance of manuscript analysis, and the challenges in authorship identification. Key concepts such as palaeography and distinctive handwriting features are also explored, along with a review of relevant datasets and previous studies.

Chapter 2:Deep Learning and Image Classification in Arabic Manuscripts
This chapter presents the theoretical foundation of deep learning techniques used in the study. It provides an overview of artificial intelligence, deep learning, and the role of Convolutional Neural Networks (CNNs) in image classification. The chapter also discusses various pre-trained CNN architectures such as VGG, ResNet, and Inception, highlighting their relevance to authorship identification tasks.

Chapter 3: Experimental Methodology and Evaluation

This chapter details the practical implementation of the project, including the dataset used, preprocessing techniques, and data augmentation methods. It presents the proposed custom CNN architecture and describes the training process, experimental setup, and evaluation metrics. The chapter also analyzes the results obtained from different models, compares their performance, and discusses the implications for authorship identification.

Conclusion

The conclusion summarizes the findings of the study and highlights its contributions to the fields of artificial intelligence and historical document analysis. We also address the challenges encountered during implementation and propose future research directions to enhance model performance and expand the dataset.

Chapter I: Identification of Ancient Manuscript Authors

Introduction

Identify Author or scribe is one of the most significant elements that can be extracted from historical manuscripts, which is essential for researching the text's identification. Numerous researchers have created automated systems that use a variety of methods to identify the writer. As deep learning technology has advanced. In this chapter, we will explain certain concepts such as manuscript, handwriting, feature extraction, and author identification, as well as the challenges of identification.

I.1 Historical Context of Manuscript Authorship:

I.1.1 Historical:

The origin of ancient manuscripts began with the early era of documentation, when diligent scholars emerged and sought to record various fields of knowledge. As the Islamic state expanded and people needed to acquire knowledge, there became a necessity for methods to preserve and disseminate it. The first sciences to spread during that period were the Hadith (Prophetic traditions) and the jurisprudence of the Companions and their followers — may Allah be pleased with them. One of the greatest features of that era was the emergence of the four prominent Imams : Imam Malik, Ahmad ibn Hanbal, Al-Shafi'i, and Abu Hanifa.[1]

The necessity to document the knowledge derived from these scholars, including, their legal opinions, and their views became more pressing. Similarly, there was a growing need for the sciences of the Arabic language to firmly support the religious sciences, leading to the documentation of grammar, morphology, rhetoric, linguistics, and mathematics .[1]

I.1.2 Manuscripts:

Manuscripts are texts written by hand on various materials such as parchment, paper, or leather before the invention of printing. The word "manuscript" is the passive participle of the verb "to write," indicating something that was written by hand. Manuscripts are considered primary sources of all information. They are original and unique works, unlike printed books, which are multiple copies of the same text. Manuscripts hold great historical and cultural value and are often preserved in libraries and museums. Some manuscripts are distinguished by their exquisite artistic decoration, including illustrations and gilded calligraphy, making them unique pieces of art. Manuscripts are written in different languages and are regarded as important historical documents[2]

Figure I.1: Manuscripts

I.2 Distinctive Handwriting Features of the Author

I.2.1 The Concept of Handwriting and Its Analysis

Handwriting is considered a highly individualized form of human expression, influenced by neuromuscular coordination and psychological factors. Through careful analysis, handwriting can be used to authenticate documents and identify individuals, as it reveals distinctive features unique to each person. Handwriting analysis relies on observing specific characteristics such as letter size, slant, pressure-t-on the paper, and consistency of movements, providing a unique ”signature” for each writer. Although environmental or health factors may sometimes affect writing patterns, the core features tend to remain relatively stable, making handwriting an effective tool in forensic and investigative fields [3].

I.2.2 Factors Influencing Handwriting Patterns (Age, Culture, Habit)

Handwriting patterns are influenced by a variety of factors, including age, cultural background, and personal habits. As individuals age, noticeable changes occur in their handwriting due to the decline in fine motor skills, leading to slower and less consistent writing. Culture plays a significant role in shaping initial writing styles, where educational and linguistic environments impact the formation and arrangement of letters. Furthermore, habitual practices developed over time create distinctive motor signatures unique to each individual, which tend to persist even under different environmental or health conditions. Understanding these influences is essential for accurately analyzing handwriting and assessing its reliability in verification processes [4].

I.2.3 Challenges in Extracting Features from Handwritten Texts

Extracting features from handwritten texts remains a challenging task due to the significant variability in writing styles, inconsistent character shapes, and the presence of noise in scanned documents. Handwriting can vary in terms of stroke thickness, pressure, and curvature between individuals, and even within an individual over time, increasing the complexity of automated processing. Furthermore, factors such as paper quality, pen type, and scanning methods can introduce additional noise into the digital images of handwriting. These challenges demand the development of robust algorithms capable of handling random variations and accurately extracting essential features to support effective handwritten text recognition [5].

I.3 The Authorship Identification Problem

I.3.1 Problem Definition

The authorship identification problem refers to the task of determining the individual who produced a given text or document based on specific characteristics present in the writing. Traditionally, this task has been crucial in forensic investigations, historical manuscript analysis, and literary studies. Authorship identification relies on the premise that each person exhibits unique, relatively stable writing patterns and stylistic choices that can be captured and analyzed. These features may include lexical usage, syntactic structures, handwriting traits (in handwritten documents), and other measurable linguistic or graphical attributes. The main challenge lies in dealing with variations in a person's writing due to different contexts, emotional states, and physical conditions, which may obscure the stable characteristics used for identification [6].

I.3.2 Applications of Authorship identification technologies:

Authorship identification technologies have broad applications across various domains. In the field of forensics, they are used to solve cases involving anonymous letters, threats, and disputed documents. In historical research, they assist scholars in attributing ancient manuscripts and anonymous works to their rightful authors. Additionally, authorship analysis has become increasingly important in cybersecurity, where identifying the author of malicious content such as phishing emails or defamatory online posts is critical. Educational institutions also employ these techniques to detect plagiarism by verifying the authorship of academic work. As technology advances, authorship identification is being integrated into digital forensics and artificial intelligence-based security systems, enhancing the ability to monitor and attribute content creation[7].

I.3.3 Traditional Methods of Authorship Identification:

Traditional approaches to authorship identification primarily relied on manual analysis by experts who examined stylistic and linguistic features in texts. These methods involved qualitative assessments, including the study of vocabulary richness, sentence structure, and handwriting traits for handwritten documents. However, traditional techniques were limited by subjectivity and scalability issues.

I.3.4 Recent Advances in Authorship Identification

In contrast, modern approaches leverage computational methods, including machine learning, natural language processing (NLP), and deep learning algorithms. These techniques automate feature extraction and analysis, allowing for the processing of large datasets with higher objectivity and accuracy. Machine learning models, such as Support Vector Machines (SVM), neural networks, and convolutional neural networks (CNNs), have significantly improved the reliability of authorship attribution, making it feasible to handle complex cases with minimal human intervention[8].

I.4 Existing Works on authorship Identification

Authorship identification is a well-established research area situated at the intersection of pattern recognition, linguistics, and forensic analysis. Traditionally, this task was performed through manual graphological analysis, where human experts examined handwriting features such as stroke shapes, letter proportions, slant, spacing, and pressure to infer the identity of a writer. Although these methods are insightful, they suffer from subjectivity and limited scalability [9]. With the rise of computational methods, numerous automated techniques have been developed, leveraging machine learning and, more recently, deep learning algorithms to identify writers based on handwriting samples. Early approaches relied on handcrafted features, including texture descriptors, contour-based features, and structural characteristics extracted from individual characters or entire lines [10]. These features were then used in conjunction with classifiers such as Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN). More recent works have shifted towards deep learning, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which learn high-level representations automatically from raw image data without requiring manual feature engineering [11]. Studies have primarily focused on Latin-script datasets such as IAM, Firemaker, and CVL, which offer labeled handwriting samples for training and evaluation. Despite the progress in writer identification, a significant limitation in the literature is the underrepresentation of non-Latin scripts, particularly Arabic.

I.5 Datasets Used for Ancient Handwriting Recognition

Datasets play a crucial role in the development and evaluation of handwriting recognition systems, particularly when dealing with ancient manuscripts. These datasets typically consist of digitized images of historical documents annotated with metadata such as the author, date, region, or script style. In the context of ancient handwriting, datasets are often limited due to the fragile condition of original documents, the scarcity of labeled data, and the linguistic diversity of historical texts [10]. Nevertheless, several datasets have been developed to facilitate research in this field.

I.5.1 The IAM Handwriting Database [9]

The IAM Handwriting Database is one of the most widely used benchmark datasets in the field of offline handwriting recognition and writer identification. Introduced by Marti and Bunke in 2002, the dataset contains forms of English handwritten text collected from 657 writers of different ages and nationalities. The text is based on the Lancaster-Oslo/Bergen (LOB) Corpus, and each writer contributed multiple pages, making the dataset rich and diverse in terms of writing styles.

It includes labeled data at the line, word, and character levels, enabling a wide range of research applications such as handwriting recognition, word spotting, writer identification, and writer verification. The dataset is particularly valuable because of its clean annotations, standardized format, and the availability of segmentation ground truth, which allows researchers to evaluate their models consistently.availability of segmentation ground truth, which allows researchers to evaluate their models consistently. The IAM dataset has served as a benchmark in many influential studies and has helped drive significant advances in the use of machine learning and deep learning in document analysis. However, as a Latin-script corpus, it does not represent the characteristics of scripts such as Arabic, Chinese, or Indic languages, which limits its applicability in research targeting diverse writing systems.

I.5.2 The Saint Gall Dataset [12]

The Saint Gall Dataset is a historical handwriting database designed specifically for the analysis of medieval manuscripts. Developed by Fischer et al. in 2011, the dataset is based on a 9th-century Latin manuscript originating from the Abbey of Saint Gall in Switzerland, which is one of the oldest surviving scriptoria in Europe. The dataset consists of 60 pages of carefully digitized and preprocessed handwritten text, containing more than 9,000 words and approximately 50,000 characters. It offers line and word-level segmentation and is commonly used for tasks such as historical handwriting recognition, writer identification, and paleographic analysis.

One of the notable aspects of the Saint Gall Dataset is its historical authenticity; it captures the unique writing styles and challenges of ancient scripts, such as irregular letter shapes, degraded ink, and inconsistent spacing. Unlike modern handwriting datasets, it reflects the real-world complexity of ancient document analysis, making it a valuable resource for developing and evaluating models in both traditional machine learning and deep learning frameworks. However, like many historical datasets, it focuses exclusively on Latin script, limiting its use for research involving non-Latin languages such as Arabic or Hebrew.

I.5.3 The CVL Dataset [13]

The CVL Dataset is a widely used handwriting database developed to support research in offline writer identification and handwriting recognition. Introduced by Kleber et al. in 2013, the dataset contains high-quality scanned images of handwritten texts provided by 310 different writers, with multiple samples per writer to account for intra-writer variability. Each writer contributed texts in both English and German, and each text is written using the same predefined content, which helps normalize comparisons between writing styles.

The dataset includes segmented lines and words, making it useful for various tasks such as writer classification, writer retrieval, and text recognition. One of the key strengths of the CVL Dataset is its diversity in handwriting styles, which enables robust training and evaluation of machine learning and deep learning models. Additionally, the consistent writing content across users facilitates fair benchmarking for handwriting-based applications. Despite its contributions, the CVL Dataset is limited to Latin scripts, and like many other widely used datasets, it lacks support for non-Latin languages, especially complex scripts like Arabic, which require further dataset development for inclusive handwriting research.

I.5.4 The KHATT Dataset [14]

The KHATT dataset, which contains modern Arabic handwriting, has also been used as a proxy for studying Arabic scripts, although it lacks the historical and stylistic richness of ancient calligraphy. The challenges of building comprehensive datasets for Arabic manuscripts include not only the complexity of the script (e.g., diacritics, ligatures, variable letter forms) but also the lack of digitized archives and author-labeled data.

As a result, researchers often rely on custom datasets, created through collaboration with libraries or cultural institutions, which involves manual digitization, labeling, and preprocessing [15]. The development of larger and more diverse datasets specific to ancient Arabic manuscripts remains a key step toward advancing deep learning models in this underexplored area of handwriting analysis.

I.5.5 The Firemaker Dataset [16]

The *Firemaker Dataset* is a specialized handwriting database primarily designed for research in writer identification and authorship analysis. It was introduced by Schomaker and Bulacu in the early 2000s as part of efforts to evaluate text-independent writer recognition methods.

The dataset contains handwritten texts produced by 250 writers, each asked to write the same Dutch text, which helps control the influence of linguistic content on handwriting features. Each sample is scanned at high resolution and segmented into lines and words, enabling detailed analysis of stylistic characteristics such as stroke curvature, slant, pressure, and pen lifts.

The Firemaker Dataset is particularly well-suited for text-independent writer identification, where the system must recognize the writer regardless of the written content. This makes it valuable for forensic document examination and biometric applications.

While the dataset is rich in stylistic variation and useful for developing and benchmarking machine learning algorithms, it is restricted to Dutch-language content and

Latin script, and does not address the complexities of non-Latin scripts such as Arabic or Chinese, which remain underrepresented in the literature.

Conclusion

The authorship identification problem presents a significant challenge in both academic research and practical applications. Most of the available studies and datasets are based on European manuscripts, and few address the complexities of Arabic calligraphy, such as ligatures, variable letter shapes, and rich stylistic diversity. This highlights the need for further research and dataset development to advance the state-of-the-art in authorship attribution for Arabic manuscripts. Modern techniques, supported by machine learning and artificial intelligence, have significantly improved the efficiency and accuracy of authorship attribution. The next chapter is devoted to explaining the machine learning and deep learning [17].

Chapter II: Deep Learning and Image Classification in Arabic Manuscripts

Introduction

Deep learning has fundamentally altered the practice of image classification in recent years. The tools available through deep learning techniques are more powerful and adaptable than traditional image classification methods. The Convolutional Neural Network (CNN) is a specific model architecture that has been particularly successful in extracting typically complicated patterns from visual data. The shift from older traditional methods of image classification to CNNs has proven to be particularly useful when trying to create models from images, in particular complex images such as historical Arabic manuscripts, which may include subtleties in the stylistics that will have value in the information being extracted. The following will touch on the fundamentals that have propelled CNNs to the forefront of contemporary image classification, as well as revealing hints of their future potential, particularly for distilling information from images that communities consider difficult to extract.

II.1 Artificial Intelligence and Deep Learning

Artificial Intelligence (AI) is a field in computer science that is concerned with the automation of intelligence and the enablement of machines to achieve complex tasks in complex environments[18].

Machine Learning (ML) is a rapidly evolving scientific domain that addresses the challenges posed by the explosion of big data and constitutes a core infrastructure for both artificial intelligence and data science[19].

Deep learning, a subfield of machine learning, aims to enhance AI systems by enabling them to learn and improve through experience and observation[20]. One of the most widely used deep learning algorithms is the Convolutional Neural Network (CNN)[30].

The advent of deep learning and CNNs marked a significant transformation in the field. These models have achieved state-of-the-art results across various tasks, even when trained on relatively small datasets. This success has increased their prominence among researchers and developers and has fueled widespread interest in both AI and ML[21].

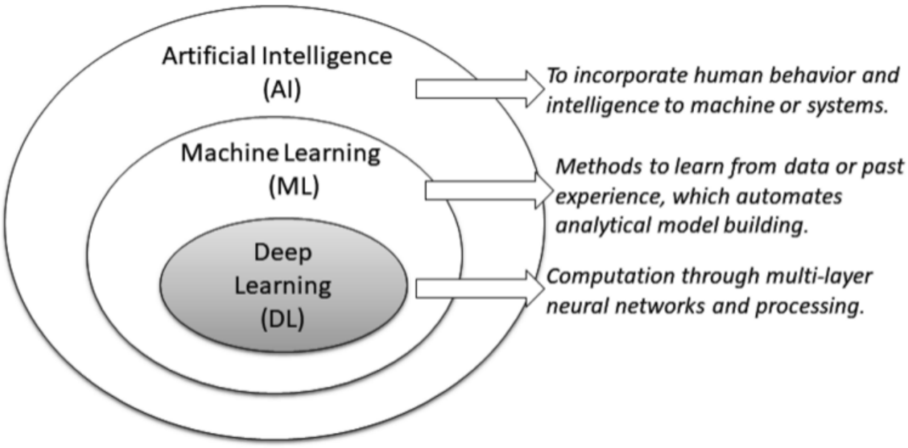


Figure II.1: An illustration of the position of DL, comparing with ML and AI[24].

- **Object detection:** Convolutional Neural Networks (CNNs) can be effectively used to detect and label multiple objects within an image, such as books, faces, or vehicles.

Figure II.2: Example of detecting objects in an image[22].

- **Image classification :** Convolutional Neural Networks (CNNs) can be effectively used to classify images into different categories, such as cats, dogs, or flowers.

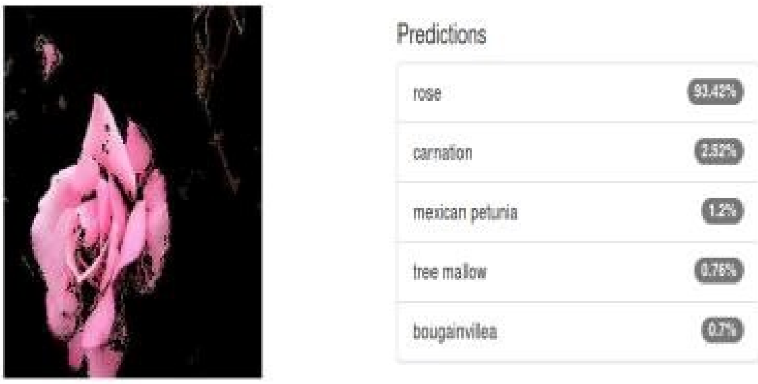


Figure II.3: Example of image classification[23].

One of the most crucial methods in deep learning is the use of Convolutional Neural Networks (CNNs), which have shown promise in several domains, including natural language processing, object recognition and tracking, and image categorization. Their significance is especially emphasized in image processing, where they are employed to identify patterns and extract visual information, which makes them appropriate for jobs like handwriting recognition, which includes identifying the authors of Arabic manuscripts by examining pictures of their writings.

II.2 Why CNN for Arabic Manuscript Author Identification:

Traditional systems of handwriting recognition have relied on handcrafted features and a large amount of prior knowledge. Training an Optical Character Recognition (OCR) system based on these prerequisites is a challenging task. Research in the handwriting recognition field is focused around deep learning techniques and has achieved breakthrough performance in the last few years. Still, the rapid growth in the amount of handwritten data and the availability of massive processing power demands improvement in recognition accuracy and deserves further investigation. Convolutional Neural Networks (CNNs) are very effective in perceiving the structure of handwritten characters/words in ways that help in automatic extraction of distinct features and make CNN the most suitable approach for solving handwriting recognition problems[24].

Some of the manuscripts created before the Hijra of the Prophet “Mohammed”—which marks the beginning of the Islamic calendar—are all handwritten and have poor visualization quality, which makes them harder to visualize and read. Sometimes the same person establishes and writes the manuscript. In other situations, one person creates the manuscript (called the “author”), while another different person writes it (called the “writer”). There might also be an “editor” that reviews the written manuscript and modifies it. However, this study considers only the “authors” of the Arabic manuscripts[25].

CNNs can differentiate between distinct writing styles of various authors based on characteristics like angles, letter curves, and word overlap since they are very good at identifying the finer details of Arabic handwriting. CNN is the perfect tool for examining old manuscripts with poor visual quality or those that have been damaged because it can continuously increase recognition accuracy by learning these qualities straight from the data.

When compared to conventional methods, CNN is especially well-suited for managing and analyzing enormous volumes of Arabic manuscript data more effectively. This offers a major benefit when working with big and complicated datasets.

Investigating the fundamental architecture of CNNs, which enables them to process data so effectively, is crucial to comprehending how neural networks can successfully differentiate between the handwriting styles of each author.

II.3 Structure of Convolutional Neural Networks (CNN)

Convolutional Neural Network (CNN), often called ConvNet, has a deep feed-forward architecture and possesses an astonishing ability to generalize more effectively compared to networks with fully connected layers[26].

CNNs can learn features from data without the need for manual feature engineering, allowing most tasks in computer vision to be performed without prior knowledge of the features. They represent a major class of deep learning algorithms used extensively for computer vision problems. This type of network utilizes a stack of convolutional blocks to extract hierarchical features from images, making it suitable for a wide variety of visual recognition tasks. Convolution is the process of weaving together two sets of information in a methodical manner. Convolutions have been widely used in image processing specifically for blurring and sharpening pictures, but many other operations as well (e.g., edge enhancement and embossing). Convolutional neural networks, create a structured method of establishing local connections between neurons within neighboring layers. They apply filters (also referred to as kernels) to identify which features like edges are present in an image. A Convolutional Neural Network (CNN) is made of a sequence of layers, each representing four basic operations:

- Convolution
- Non Linearity (ReLU)
- Pooling or Sub Sampling
- Classification (Fully connected layer)

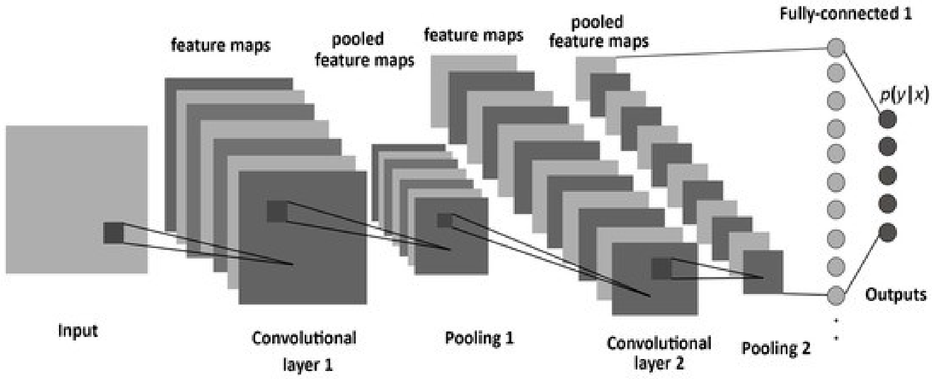


Figure II.4: The structure of a CNN[27].

II.3.1 Convolution Operation

The first layer of a Convolutional Neural Network is always a **Convolutional Layer**. Convolutional layers take the input and apply a convolution to the input and then pass the result to the following layer. A convolution converts all the pixels in the receptive field into a single value. For example, if we apply a convolution to an image, we shrink the size of the original image while merging all the information in the receptive field into a single pixel. The eventual result of the convolutional layer is a vector. Depending on the problem we need to solve and the type of features we are trying to learn, there are varying types of convolutions available for use.

The **2D convolution layer**, abbreviated as `conv2D`, is the most often used form of convolution. In a conv2D layer, a filter or kernel "slides" through the 2D input data, executing element-wise multiplication. As a consequence, the results are summed into a single output pixel. For each point it slides over, the kernel performs the same procedure, transforming a 2D matrix (image) into a distinct 2D matrix of features[20]. When working with 3D images—for instance, images with depth values or multiple sub-components through the color channels—we can use a **3D kernel** ($n \times n \times n$), where the receptive field becomes a 3D grid and image units are referred to as *voxels*.

The filter size plays a critical role in feature extraction: **larger filters** are useful for capturing global features (such as ink distribution patterns or stroke densities over large areas of manuscripts), while **smaller filters** are more effective for capturing fine local details (such as curves of letters, types of pen pressure, and edge details). These local features are often essential for distinguishing handwriting styles between different authors.

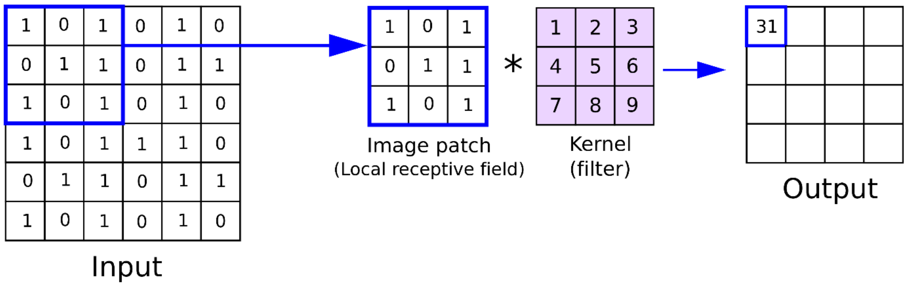


Figure II.5: Convolutional layer[29].

II.3.2 Non Linearity

Once the feature maps have been created, the next step is to apply a non-linear activation function. An activation function plays an essential role in CNN layers[30]. Simply put, this step is crucial as it improves neural networks and shortens training time. It helps the network learn and fit more complex relationships between input and output data, which means that the accuracy errors of the neural network can be reduced due to faster convergence compared to traditional approaches.

One of the simplest gradient computations depends on the sign of x, whether at $x = 0$ or $x = 1$. The ReLU function is defined as:

$$\text{ReLU}(x) = \max(0, x)$$

However, this approach ignores negative values, which may hold some useful information. In some cases, small negative values could be useful. This is why a new activation function is proposed, called Leaky ReLU, defined as:

$$\text{Leaky ReLU}(x) = \max(0, x) + \alpha \cdot \min(0, x)$$

where α is a small pre-determined parameter (typically $\alpha = 0.01$) that can also be learned.

II.3.3 Max Pooling

In this phase, we replace an $n \times n$ area with the maximum value within that region. This operation serves two main purposes:

- It selects the maximum activation in a local region, thus providing a small form of spatial invariance.
- It reduces the size of the activation maps for the next layer by a factor of n^2, and as a result, fewer parameters need to be learned in the subsequent layers.

There are other types of polling such as global-average-pooling, winner-takes-all-pooling, and stochastic-pooling.

Figure II.6: example of max-pooling[31].

II.3.4 Fully connected layer

The final phase in the CNN architecture is the fully connected layer or a sub neural networks that has the right to decide or make segmentation after getting features from the previous convolution block.

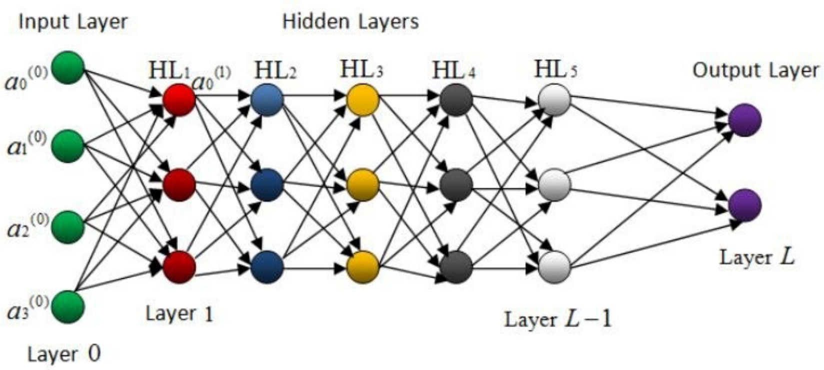


Figure II.7: A fully connected layer[31].

II.4 Pre-trained CNN Architectures

Pre-trained models are an incredible asset to deep learning, with computer vision being the most notable. Pre-trained models allow you to quickly and easily solve a new problem on less data and with limited computations by using models that were trained on massive datasets like ImageNet. Pre-trained models are to prevent you from having to start training the model from scratch, and this saves you time and costs in achieving good performance. Most pre-trained models are based on Convolutional Neural Networks (CNNs), which have the best track record for image analysis and visual pattern recognition. There are many reputed CNN architectures. The description of most popular CNN architectures is given below.

II.4.1 Visual Geometry Group (VGG):

VGG is a traditional CNN architecture, created to increase the depth of CNNs by using small 3×3 filters and containing pooling layers and a fully connected layer. VGG was introduced first by two Oxford researchers from the Visual Geometry Group lab in 2014 and has been influential in the computer vision space because it achieved state-of-the-art performance on several classification and recognition benchmark datasets, including the ImageNet Large Scale Visual Recognition Challenge (ILSVRC).

The VGG network has six different CNN configurations, which are VGG11, VGG11 (LRN), VGG13, VGG16 (Conv1), VGG16, and VGG19. The convolution layers in the model are represented by the numbers “11”, “13”, “16”, and “19”[32].

Although VGG19 has three additional layers, which increases its complexity and resource consumption, VGG16 is considered more efficient due to having fewer parameters.

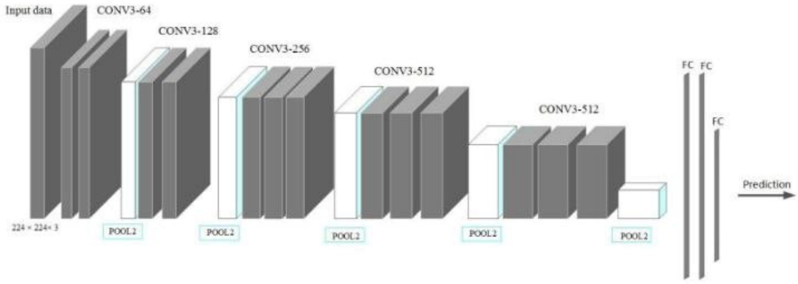


Figure II.8: VGG Architecture[30].

II.4.2 Residual Neural Network (ResNet):

ResNet is a significant deep learning model where weight layers learn residual functions with respect to the layer inputs. ResNet was developed in 2015 for image recognition and excelled in the ImageNet challenge. It is based on different types of residual blocks, including basic blocks that consist of two 3×3 convolutional layers with a residual connection, bottleneck blocks that have three sequential convolutional layers for dimension reduction and restoration, and pre-activation blocks to reduce non-identity mappings between blocks.

It addresses the vanishing/exploding gradient problem by utilizing identity skip connections that allow the gradient to flow along an extra path (shortcut path). There are different variants of ResNet, such as ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152[33].

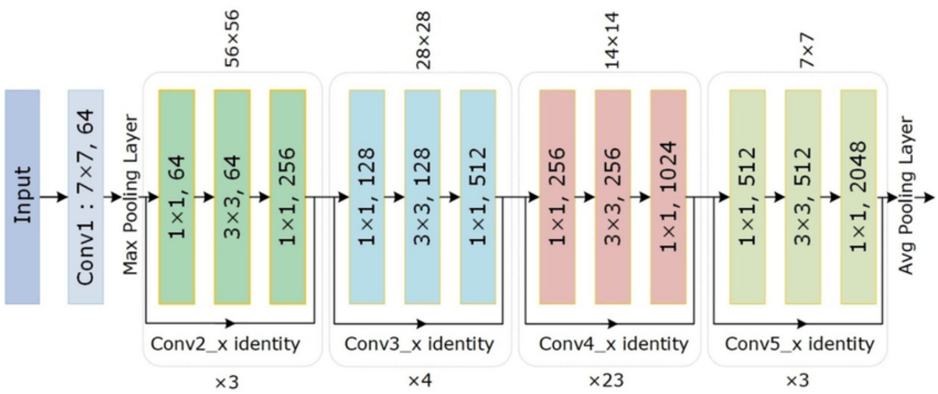
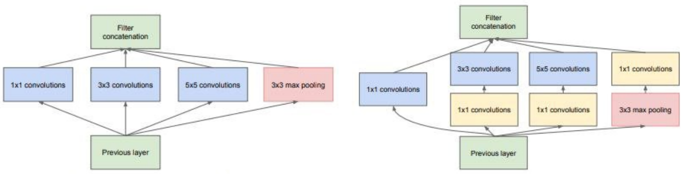


Figure II.9: The general architecture of ResNet.

[34]

II.4.3 Inception Architecture (GoogLeNet)

The Inception architecture, also known as GoogLeNet, was introduced by Szegedy et al. in 2014 and marked a significant step forward in efficiency for deep convolutional neural networks. These types of convolutions effectively make the model deeper and cheaper to compute. One of the key contributions of the Inception architecture is to increase the number of convolutions happening in parallel, and increase the depth of the network, without exceeding a computational budget. GoogLeNet, or the first version of the Inception architecture, was composed of 22 layers and performed very well in the ILSVRC 2014 (ImageNet Large Scale Visual Recognition Challenge). Inception-v2, Inception-v3, and Inception-v4, or additional versions of the Inception architecture, incorporate extra improvements such as batch normalization, factorization of larger convolutions into smaller convolutions, and residual connections to increase model accuracy and training consistency.



(a) Inception module, naïve version (b) Inception module with dimension reductions

Figure II.10: Inception module.

[35]

Advanced architectural approaches, such as CNNs, and their modified versions, VGG-16, ResNet, Inception, significantly contribute to improving the performance of Arabic manuscript author identification systems by extracting detailed and accurate patterns from images, including curves and overlaps in letters, allowing for better separation of author handwriting features. When investigating deep learning networks' application to Arabic manuscripts, the deep learning model responds to various aspects of uniqueness, including lines, letters' connectivity to one another, and uncertainty in the recognition of letters due to unclear or damaged handwriting. The deep learning models exhibited remarkable capabilities in recognizing barely legible letters, and, by adopting a pre-trained architecture trained on large datasets like ImageNet, training time was significantly reduced and required less data, elucidating relationships to facilitate more effective author identification of Arabic manuscript authors. Hence, the variation of the individual architectures suggests a degree of important functionality reliant on assistance of identifying handwriting traits even with poor quality manuscripts containing unclear elements, as well as providing more opportunities for a wider range of manuscript analysis research activities.

Conclusion

In this chapter, we have discussed the basic concepts of image classification, deep learning, and highlighted architectures that have defined the field. We have described the important role of convolutional neural networks and the structural elements of these networks, and how these architectures assist in improved performance in tasks such as author identification for Arabic manuscripts. This foundational knowledge will help frame discussions surrounding the many complex concepts presented in the following chapter.

Chapter III: Experimental Methodology and Evaluation

Introduction

In this chapter, we present the practical phase of our project, detailing the steps undertaken to develop an effective model for classifying manuscript images using Convolutional Neural Networks (CNNs). The process begins with dataset preparation, including preprocessing and data augmentation techniques, followed by model construction and experimentation with several pretrained architectures. The focus is specifically placed on Arabic manuscripts, aiming to enhance classification performance within this underexplored domain.

III.1 Dataset

Due to the lack of readily available and complete information about the authors of historical Arabic manuscripts in the form of a reliable dataset, we were compelled to collect the data manually. The first challenge we faced was the difficulty of finding manuscript images that could be reliably attributed to specific authors. The image data was often scattered and lacked organization, which made classification and sorting particularly difficult. Initially, we were only able to find 3 to 4 images per author—an insufficient quantity for training deep learning models, which typically require large and diverse datasets. We continued our search using digital libraries and institutional archives, most notably Alreq AlManshour platform for manuscript cataloging [38], the "Al-Kitabdar" Islamic Manuscript Directory [39], and the ILM Foundation for Heritage and Digital Services [40]. Eventually, we managed to collect around 20 images per author. Although this number remains relatively limited given our goal of building an effective model, it represented a significant improvement compared to the initial batch of images. We manually verified that the collected images were consistent in terms of visual or textual style specific to each author, to provide the model with a better chance of detecting variations between different handwriting styles. Nonetheless, we encountered other challenges, such as the poor quality of some images and the inaccurate labeling of certain manuscripts, which required manual verification of each image. We organized the images into folders, with each folder containing images for a single author, to ensure category separation during training. This organization was based on the assumption that each author has a unique handwriting style that can be identified through features such as letter shapes, spacing between characters, placement of diacritical marks, and pen stroke thickness. In the end, we constructed a dataset consisting of 10 distinct classes (each representing a different author), with approximately 20 images per class. The images varied in terms of source type: some depicted full manuscript pages, while others were cropped excerpts that revealed clearer handwriting characteristics. This diversity enabled the model to better learn and distinguish the writing patterns associated with each author.

Figure III.1: Dataset

III.2 Pre-Processing

At this stage of the project, we prepared the Arabic manuscript images for processing in a deep learning environment. As the images originated from a variety of sources, there was considerable variation in terms of size, quality, lighting conditions, and ink thickness. These inconsistencies necessitated a systematic and consistent preprocessing pipeline to ensure that differences in image appearance did not hinder the model's ability to recognize writing patterns. Proper preprocessing was also essential to avoid misinterpreting similar Arabic scripts as distinct samples and to enhance overall model performance. The first preprocessing step involved resizing all images to a fixed dimension of 224×224 pixels, aligning with common input requirements for convolutional neural network (CNN) architectures. This resizing step not only standardized the input data but also facilitated compatibility with pretrained models and optimized the training process. Subsequently, we normalized pixel values by converting the original intensity range of [0, 255] to a [0, 1] scale. This was achieved by converting the image

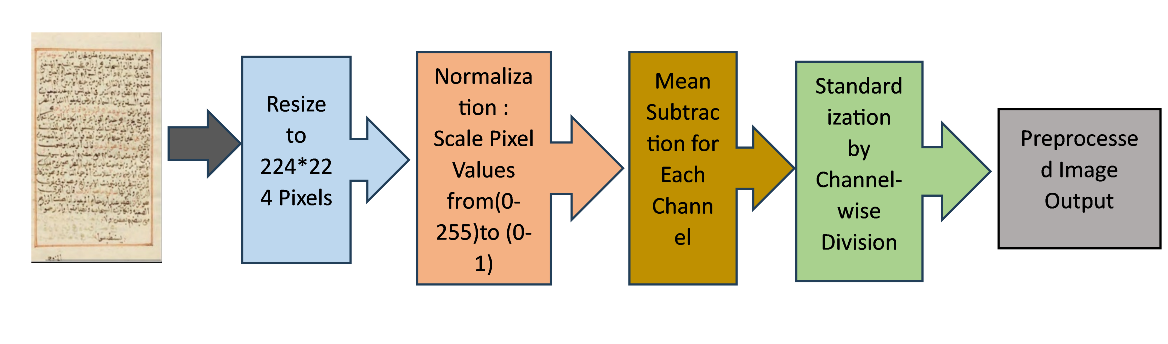


Figure III.2: Pre-Processing Pipeline for Manuscript Images

data type to float32 and dividing each pixel value by 255. This normalization reduced the overall variance in pixel intensities and allowed the model to focus on subtle handwriting characteristics—such as the curvature of letters, the connections between characters, and variations in stroke direction—rather than being distracted by large differences in brightness or paper texture. To further reduce the impact of irrelevant visual information, we applied mean subtraction on the RGB channels using the ImageNet means: [0.485, 0.456, 0.406]. This step helped neutralize variations caused by different ink tones or paper backgrounds, allowing the model to center its learning on meaningful features of Arabic script structure, rather than being influenced by color or lighting inconsistencies. Following this, each image was standardized by dividing the pixel values by their respective standard deviations: [0.229, 0.224, 0.225]. This normalization ensured consistent contrast and intensity across the dataset. Given the variability in ink density and paper absorption across manuscripts, this step ensured the model focused on structural features—such as shape, alignment, and spacing—rather than superficial intensity cues. Although normalization parameters like these are typically used with models pretrained on ImageNet, we applied the same preprocessing steps to our own CNN model. The goal was to provide a stabilizing effect during training and enhance model accuracy, especially since we had limited control over the physical and scanning conditions of the source documents.

III.3 Data-augmentation

Given the limited number of images in our dataset—no more than 20 images per class—there was a clear constraint on the ability of a deep learning model to generalize effectively. This scarcity of training data posed a significant risk of overfitting, as the model could easily memorize the small set of samples without learning meaningful patterns applicable to unseen data. To address this challenge, we implemented data augmentation techniques. Data augmentation provided a practical and effective solution by artificially expanding the dataset and introducing controlled variations that simulate real-world diversity in manuscript appearances. This step was essential not only to mitigate overfitting but also to enhance the model's ability to accommodate the natural complexity and significant variability inherent in Arabic handwriting. By applying transformations such as rotation, scaling, flipping, brightness adjustment, and noise injection, we aimed to replicate realistic conditions under which historical manuscripts may have been written, scanned, or preserved. These transformations enriched the dataset and helped the model become more robust to inconsistencies in document conditions, ultimately contributing to improved classification performance. We used the KerasImageDataGenerator class to apply a suite of transformations to the original image data as follows:

- rotation-range: We made rotations of our images in the range of ±20% degrees. Such changes represent changes in writing angles of individuals or even variances created by the scanner.

- zoom_range: At the ±15 % zoom level, we used zooming to simulate distance from and/or diversity of focus in the process of digitizing documents, providing additional visual diversity in the data set.

- width-shift-range and height-shift-range: A horizontal and vertical shift of 20 % was used to simulate shifts associated with either misalignment or cropping during image capture or scanning.

- shear-range: A light shear was used to simulate distortions in old/aged documents.

- horizontal-flip: In some cases, the images Wwere flipped horizontally depending on the manuscript. The horizontal flip was allowed because it contributed to structural diversity in the image.

- fill-mode='nearest': This mode filled new pixels with the value only taken from the nearest pixel to keep the pixels ordered after being transformed. This provided some continuity in the visuals. Therefore, every image was able to preserve identity and structural presence after applying the transformations. These processes provided us with an opportunity to expand the size and diversity of the Arabic manuscripts dataset, overcoming the problem posed by the original limited number of images

All of these transformations were only applied for the training set (80%); the validation set and test set (10% each) were left unchanged to evaluate the model performance under as objective on new data

III.4 Technologies and Environment

For our CNN architecture and data pre-process we used Python as the main programming language, we also used TensorFlow and its sub-libraries as well as several other matrix manipulation libraries. This research was conducted on Google Colab, utilizing the Tesla T4 GPU with 15GB of GPU memory and approximately 12.6GB of RAM. The data was stored and accessed from Google Drive for ease of management and access during the training process

III.5 Proposed Architecture

The model used for this experiment is a Convolutional Neural Network (CNN) which is implemented by an open-source deep learning library known as Keras. Keras is a library which allows using CNN architecture for image classification quite easily. A CNN is comprised of many layers, each has a different function with respect to processing the information that is input into it. These are all layers of the model architecture:

Model Architecture

1. **Input Layer:**

    - **Input shape:** (224, 224, 3) height, width, and number of channels since they are all RGB color images.

- • This shape details the dimensions of the images present in the dataset

2. **Convolutional Layers:**

   - `Layer conv2d`: has 16 filters, shape (3x3), with ReLU activation.
   - `Layer conv2d_1`: has 32 filters, shape (3x3), with ReLU activation.
   - The convolutional layers learn semi-coarse and coarse visual feature representations in the input images based on filters applied acrossthe images

3. **Max PoolingLayers:**

   - `Layer max_pooling2d:` : This layer follows the first convolutional layer, with a (2x2) pooling window.
   - `Layer max_pooling2d_1`: This layer follows the second convolutional layer, with a (2x2) pooling window.
   - The purpose of the MaxPooling layer is to reduce

4. **Flatten Layer:**

   - `Layer flatten`: This layer reshapes the 2-dimensional feature maps into 1-dimensional vectors to reshape the data appropriately for the dense layers

5. **Dense Layer:**

   - `Layer dense`:This layer is a fully connected layer with 64 units and ReLU layer activation.
   - This layer will learn a high level representation of the features based on the output of the convolutional and pooling layers.

6. **Dropout Layer :**

   - `Layer dropout`:This is after dense_2 and has a dropout rate of 0.3
   - This layer will randomly drop some of the layers when training as a form of combatting overfitting and increasing the generalization of unseen inputs.

7. **Output Layer :**

- `Layer dense_1`:This is the fully connected output layer and has a number of units equal to num_classes and Softmax activation

8. **Compilation and Training**

- The model was compiled utilizing the Adam optimizer, which is an adaptive learning rate optimization algorithm.
- The loss function was categorical cross-entropy which defines a loss suitable for multi-class classification.
- Accuracy was the performance measure used to monitor model performance, during training.
- The model was trained on 32 batch sizes for 50 epochs.
- Early Stopping was defined to monitor the loss improvement and stop training when no improvement occurred for a defined amount of epochs, but was not activated in this experiment. The architecture learns multi-level features by going through multiple conv and pooling layers on the input images, while the dense layers represent high-level representations of the learned features to amplify classification accuracy. Dropout is used to reduce overfitting and improve general

Table III.1: Model Summary cnn

Layer (type)	**Output Shape**	**Param #**
Conv2D (conv2d)	(None, 222, 222, 16)	448
MaxPooling2D (max_pooling2d)	(None, 111, 111, 16)	0
Conv2D (conv2d_1)	(None, 109, 109, 32)	4,640
MaxPooling2D (max_pooling2d_1)	(None, 54, 54, 32)	0
Flatten (flatten)	(None, 93312)	0
Dense (dense)	(None, 64)	5,972,032
Dropout (dropout)	(None, 64)	0
Dense (dense_1)	(None, 10)	650

Total params :5,977,770 (22.80 MB)

III.6 Transfer Learning Architectures

Besides the proposed architecture based on a type of conventional Convolutional Neural Network (CNN), this work also implement transfer learning algorithms using deep models previously trained on ImageNet i.e., VGG16, ResNet50, and InceptionV3. The aim of this approach is not only to improve classification performance, but also to extract high-level visual features more efficiently with a more limited quantity of manuscript image data. These models are key for automatic identification of Arabic manuscript authors, because they learn fine and abstract patterns within an image. This makes a dramatic difference to discriminative power, compared to constructing a CNN from scratch.

1. **VGG16:** The VGG16 model was used as the base model in a transfer learning approach. It was loaded with weights from the ImageNet dataset, and set to not include the original classification layers (include_top=False). To keep the visual features already learned, all of the base model layers were frozen. To adapt the model to manuscript authorship classification (10 classes) a custom classification head was added with:

   - A Flatten layer
   - A Dense layer with 256 units and ReLU activation
   - A Dropout layer at 0.3
   - A final Dense layer with softmax activation depending on the number of classes

   This model was compiled using the Adam optimizer with a learning rate of 1e-4, categorical crossentropy loss function and trained at 15 epochs with a batch size of 32 with image normalization based on ImageNet preprocessing

Table III.2: Model Summary VGG-16

Layer (type)	**Output Shape**	**Param #**
VGG16 (Functional)	(None, 7, 7, 512)	14,714,688
Flatten (flatten_1)	(None, 25088)	0
Dense (dense_2)	(None, 256)	6,422,784
Dropout (dropout_1)	(None, 256)	0
Dense (dense_3)	(None, 10)	2,570

Total params :21,140,042 (80.64 MB)

Trainable params:6,425,354 (24.51 MB)

Non-trainable params:14,714,688 (56.13 MB)

2. **InceptionV3:** The InceptionV3 model was used as the base network in a transfer learning context and loaded with its weights pre-trained on the ImageNet dataset (with include_top=False, I only loaded weights from the base network, not the top classification layers). The model layers were frozen to retain the pre-trained visual features. To customize the model for the Arabic manuscript author classification task (10 classes), a custom classification head was built made up of:

   - A GlobalAveragePooling2D layer to reduce spatial dimensions.
   - A Dropout layer with a rate of 0.5 to alleviate overfitting.
   - A Dense layer with 128 units and ReLU activation.
   - Another Dropout layer with a rate of 0.3.
   - A Dense layer with a softmax activation corresponding to the number of classes.

   The model was compiled using the Adam optimizer at a learning rate of 1e3, and the model was trained using categorical cross entropy loss for 15 epochs with a batch size of 32. The input images were normalized according to the standards imposed by ImageNet

Table III.3: Model Summary InceptionV3

Layer (type)	**Output Shape**	**Param #**
InceptionV3 (Functional)	(None, 8, 8, 2048)	21,802,784
GlobalAveragePooling2D (global_average_pooling2d_3)	(None, 2048)	0
Dropout (dropout_6)	(None, 2048)	0
Dense (dense_6)	(None, 128)	262,272
Dropout (dropout_7)	(None, 128)	0
Dense (dense_7)	(None, 10)	1,290

Total params:22,066,346 (84.18 MB)

Trainable params :263,562 (1.01 MB)

Non-trainable params :21,802,784 (83.17 MB)

3. **ResNet50:** The ResNet50 model was used as the base architecture for transfer learning, using pre-trained weights on the ImageNet dataset (without the original top classification layers using include_top=False) and global average pooling (pooling='avg'). All of the layers in the ResNet50 base model were frozen to maintain the pre-learned visual representations. To modify the model for the Arabic manuscript authorship classification task (10 classes) a custom classification head was introduced as follows:

   - A Dropout layer at a rate of 0.3
   - A Dense layer of 256 units, and ReLU activation
   - A final Dense layer with softmax activation to match the number of target classes.

   The model was compiled with the Adam optimizer at a learning rate of 1e-5, and trained with categorical cross entropy loss function. Training took place for 30 epochs with a batch size of 32 and images were normalized following a method compatible with ImageNet statistics

Table III.4: Model Summary ResNet50

Layer (type)	**Output Shape**	**Param #**
ResNet50 (Functional)	(None, 2048)	23,587,712
Dropout	(None, 2048)	0
Dense	(None, 256)	524,544
Dense_1	(None, 10)	2,570

Total params:24,114,826 (91.99 MB)

Trainable params:527,114 (2.01 MB)

Non-trainable params :23,587,712 (89.98 MB)

III.7 Experimental Results

This section briefly presents the experimental results obtained from applying various deep learning models to the task of Arabic manuscript authorship classification.

III.7.1 Evaluation Metrics

To properly assess the model's performance in the task of Arabic manuscript author classification, a set of both quantitative and visual metrics was employed. The primary metric used is accuracy, which refers to the proportion of correctly classified images out of the total number of images. This provides an overall view of how well the model is performing. Accuracy is calculated using the following formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{III.1}$$

- **True Positives (TP)**: Cases where the model correctly identified the actual author of the manuscript.
- **True Negatives (TN)**: Cases where the model correctly recognized that the manuscript does not belong to a certain author.
- **False Positives (FP)**: Cases where the model incorrectly assigned a manuscript to the wrong author.

- **False Negatives (FN)**: Cases where the model failed to identify a manuscript that actually belongs to a certain author.

In addition to accuracy, several other metrics were used to provide a more detailed evaluation of the model's behavior:

- **Precision:** Measures how reliable the model is when predicting a specific author. A high precision value indicates that most of the manuscripts predicted to belong to a certain author are indeed correct.

The equation for Precision is given by:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{III.2}$$

- **Recall:** Evaluates the model's ability to identify all manuscripts that truly belong to a specific author. A high recall means the model successfully retrieved most of the relevant manuscripts. The equation for Recall is defined as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{III.3}$$

- **F1-Score:** A balanced metric that combines both precision and recall. It is especially useful when dealing with imbalanced data. The equation for F1-score is given by:

$$\text{F1-Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{III.4}$$

These values were derived from the classification report, which presents the precision, recall, and F1-score for each author class. In addition, macro and weighted averages were included to provide a more comprehensive view of model performance across all classes. A confusion matrix was also used to visually summarize how the model performed for each class. It displays the predicted classes versus the actual ones, helping to identify which classes the model often confuses. This is especially relevant in Arabic manuscripts, where stylistic similarities between authors may lead to misclassifications. Finally, training and validation curves were used to monitor how the model's learning progressed over time. These plots display the evolution of loss and accuracy across training epochs and help determine whether the model is:

- **Overfitting:** Overfitting occurs when a machine learning model captures not only the underlying patterns in the training data but also the noise and outliers, resulting in excellent performance on the training dataset but poor generalization to unseen data. This phenomenon indicates that the model has learned the training data too well, including its idiosyncrasies, and therefore fails to perform reliably on validation or test sets[36].

- **Underfitting :** Underfitting refers to a situation where a model is too simplistic to learn the underlying structure of the data, leading to poor performance on both training and validation datasets. It often results from insufficient model complexity, inadequate training, or inappropriate feature selection, and indicates that the model has not captured the essential patterns necessary for accurate predictions[36].

- **Effective learning :** Effective learning in machine learning describes a model's ability to incrementally improve its performance as it is exposed to more data or training iterations. It is characterized by a gradual reduction in training and validation error, indicating that the model is successfully capturing relevant patterns in the data and generalizing well to unseen samples[37].

III.7.2 Results per Model

1. **Performance Analysis of the Custom CNN Model :** The model successfully classified manuscript authors and yielded an overall accuracy classification score of 90%, which suggests that it was successful in this regard. The measurements of Precision (0.85), Recall (0.90), and F1-score (0.8667), further suggest that the model performed equally well across all the author classifications and exhibited little tendency or bias toward any of the classes on average. This similarity between the measurements is a reflection of the models ability to recognize and differentiate the similarities and differences in patterns in the authors' data representing a high quality representation and learning was achieved while training the model.

    - **Classification Report**

Table III.5: Classificcation Raport of the CNN Model

Class (Author)	**Precision**	**Recall**	**F1-Score**	**Support**
Abu_al_Hassan_abdelFattah_Al_Mahmoudi_Al_ladhqi	1.00	1.00	1.00	2
Al_Allamah Yusuf ibn Abdullah Al-Namari Al-Qurtubi	1.00	1.00	1.00	2
Al-Imam Abu Bakr Al-Turtushi	1.00	1.00	1.00	2
Al-Imam Al-Nawawi	1.00	1.00	1.00	2
Al-Imam Al-Sanusi Al-Shafi'i	1.00	1.00	1.00	2
Al-Imam Ibn al-Qayyim al-Jawziyyah	1.00	1.00	1.00	2
Ali_ibn_Mohammad_ibn_Abd_al-Samad_al-Sakhawi_al-Dimashqi	1.00	1.00	1.00	2
Hocin_bacha_siri_najl_almrfor_laho_ibrahim	0.00	0.00	0.00	2
Iskandar_jamal	0.50	1.00	0.67	2
Mobaka_om_mohmmed	1.00	1.00	1.00	2

Overall, the classification report shows that the model perfectly classified most of the author classes (precision, recall, and F1-score of 1.00) while not performing as well on two classes: hocin_bacha_siri_najl_almrfor_laho_ibrahimand iskandar_jamal. The model there confused the other class, most likely owing to significant intra class variability, low distinctiveness, or possibly poor feature learning because it was not able to classify this class correctly. The model did recall all of the samples of iskandar_jamal, although with a quality of precision indicating that it was possibly confused with other classes.

- **Confusion Matrix**

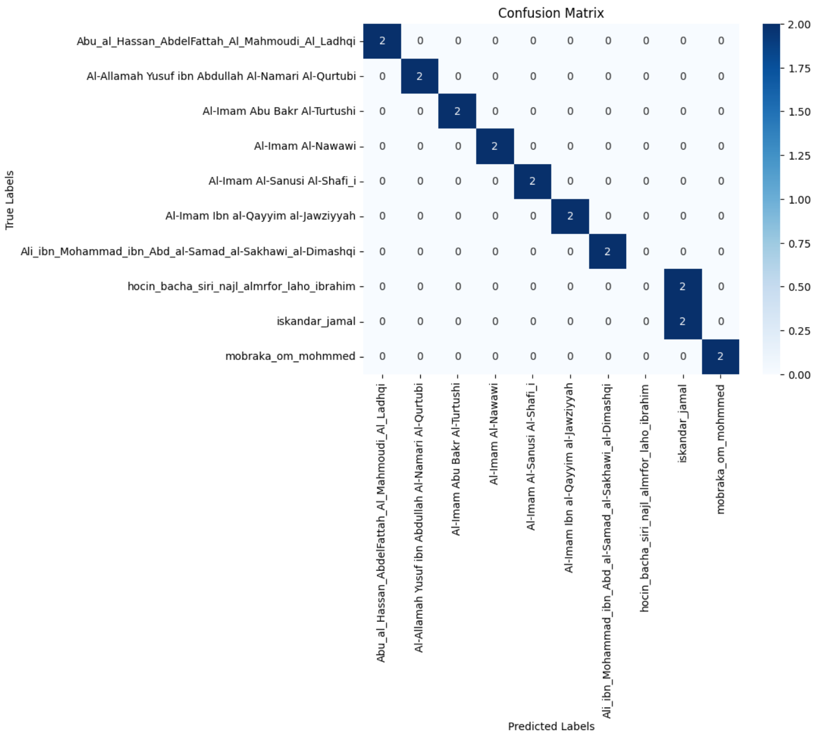


Figure III.3: Confusion Matrix of the Cnn Model

Overall, the confusion matrix appears to align with the results from the classification report, indicating that the model classified most of the author classes correctly but had difficulty correctly classifying certain samples. Specifically, the model fully misclassified one author's samples as those of another: all samples from the class hocin_bacha_siri_najl_almrfor_laho_ibrahim were incorrectly labeled as iskandar_jamal. Furthermore, some aspects of the confusion matrix help explain why the recall for the class iskandar_jamal was perfect, while some samples from other authors were classified as iskandar_jamal. Since these misclassifications came from non-iskandar_jamal classes into iskandar_jamal, they contributed to false positives, which in turn lowered the precision for iskandar_jamal.

- **Training and Validation Accuracy and Loss Plots** The accuracy curve shows that the model's accuracy performance continues to improve continuously, a little slowly and surely, until it approximately reaches 90% accuracy. This evidence shows that the model has learned from the data. The validation accuracy is about the same as the training accuracy. There are fluctuations in both training and validation accuracies, but they are minor differences something that would be expected. The similarities indicate the model generalized well and did not overfit. In other words, it is NOT just "memorizing" training data; it is learning a process to make predictions on test data that it has never encountered.

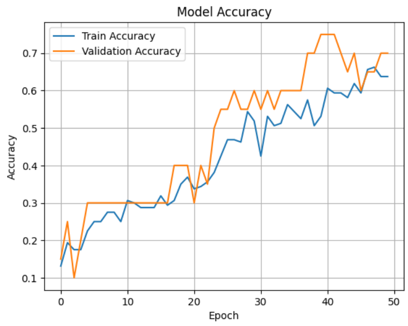


Figure III.4: Training and Validation Accuracy of the CNN Model

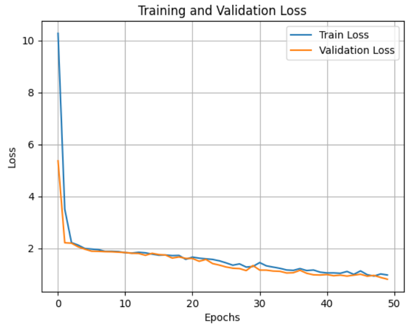


Figure III.5: Training and Validation Loss of the CNN Model

The loss curve shows that the training loss continues to clearly and steadily decline. This indicates that the errror is declining while the model is in the learning phase. The validation loss reasonably and steadily declines, there is some fluctuation in loss at times, although with limited data that is expected. The important part is, there were NO sharp or unexpected spikes in loss which indicates that the model did not overfit and is not underfit. These two observations suggest that the model was capable of learning a process, while also demonstrating stability while validating.

2. **Performance Analysis of the VGG16 Model :** The VGG16 model was able to classify manuscript authors successfully and had strong overall classification performance with an overall classification accuracy of 95%. Additionally, the precision (0.9667), recall (0.95), and F1-score (0.9467) evaluations show that the model performed similarly across the various authors. Because these values are closely aligned it can be inferred that the model was able to leverage distinct features of each author's writing style and did not exhibit any substantial bias towards any author classes. The uniformity of these evaluations implies the model was able to generalize properly, differentiate similar features, and maintain a quality representation of the learned features while training.

- **Classification Report** The classification report of VGG16 showed that accuracy levels are high, and the model classified most manuscript author categories correctly. In fact, with the direct class-specific data set used for supporting, VGG16 reached perfect (1.00) precision, recall, and F1-score in most of the author categories. This suggests that the model can identify the three different author styles and patterns of writing. However, some classes presented slightly less successful performance:
  Abu_al_Hassan_AbdelFattah_Al_Mahmoudi_Al_Ladhqi: The VGG16 model achieved a precision of 0.6667 and perfect recall of 1.0000 (meaning it retrieved all samples), but the model was not always correct; it confused it with another class. Al-Imam Al-Sanusi Al-Shafi_i: The recall for this class was 0.5000, while the precision was perfect, indicating that the model classified one of the two samples correctly, and the other one wrongly. Overall, VGG16 showed very strong performance to trust with clear ability to identify distinct stylistic features of each author, if only some minor errors in some specific classes. This suggests effective training and strong understanding of the features presented in this data.

Table III.6: Classification Report of the VGG Model

Class	Precision	Recall	F1-score	Support
Abu_al_Hassan_abdelFattah_Al_Mahmoudi_Al_ladhqi	0.6667	1.00	0.8000	2
Al_Allamah Yusuf ibn Abdullah Al-Namari Al-Qurtubi	1.00	1.00	1.00	2
Al-Imam Abu Bakr Al-Turtushi	1.00	1.00	1.00	2
Al-Imam Al-Nawawi	1.00	1.00	1.00	2
Al-Imam Al-Sanusi Al-Shafi_i	1.00	0.50	0.6667	2
Al-Imam Ibn al-Qayyim al-Jawziyyah	1.00	1.00	1.00	2
Ali_ibn_Mohammad_ibn_Abd_al-Samad_al-Sakhawi_al-Dimashqi	1.00	1.00	1.00	2
Hocin_bacha_siri_najl_almrfor_laho_ibrahim	1.00	1.00	1.00	2
Iskandar_jamal	1.00	1.00	0.67	2
Mobraka_om_mohmmed	1.00	1.00	1.00	2

- **Confusion Matrix of the VGG-16 Model**

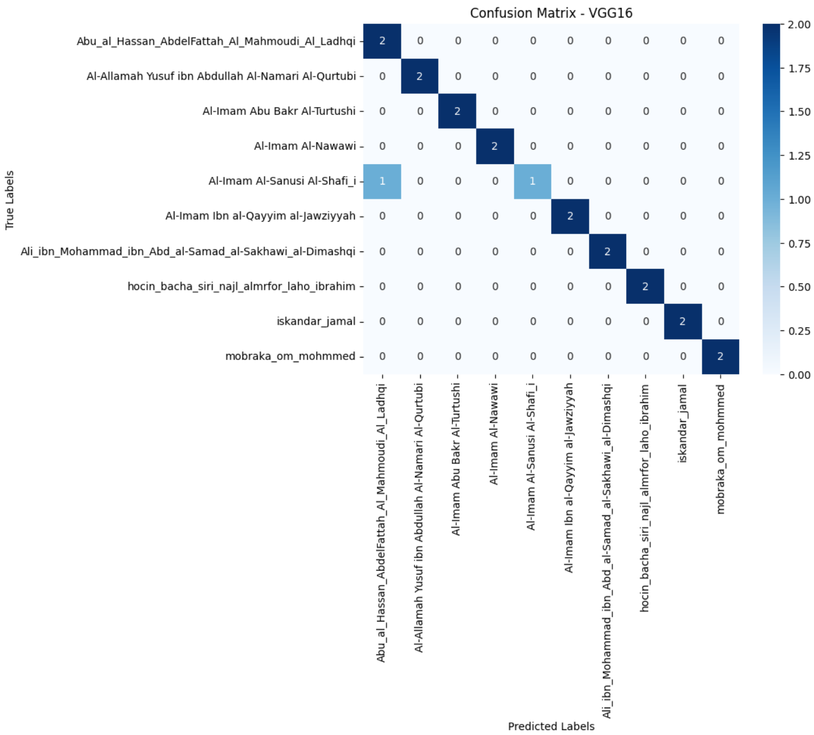


Figure III.6: Confusion Matrix of the VGG Model

Overall, the confusion matrix agrees with the classification report and supports additional detail related to specific misclassifications made by the VGG16 model. Specifically, the model correctly classified most of the authors' classes with some misclassifications. For example, one of Al-Imam Al-Sanusi AlShafi_i's samples was misclassified as Abu_al_Hassan_AbdelFattah _Al_Mahmoudi_Al_Ladhqi. This misclassification provides some insight into the lower recall (0.50) we received on Al-Sanusi, because only one of the two samples was positively (correctly) classified. Likewise, this error also affected the precision of Abu_al_Hassan_ AbdelFattah_Al_Mahmoudi_Al_Ladhqito a score of 0.6667 because one of the two samples predicted as his actually belonged to Al-Sanusi. These illustrations reflect how a single sample can incorrectly influence both precision and recall. All other author classes were perfectly classified, which shows that the VGG16 model was capable of distinguishing amongst most classes highly confidently and with accuracy, as supported by the macro-averaged metrics (Precision: 0.9667, Recall: 0.95, F1-score: 0.9467).

- Training and Validation Accuracy and Loss Plot The training and validation accuracy curves show a consistent and fair improvement in the VGG16 model's performance. Training accuracy increased gradually from the first epochs and exceeded 90% in the later epochs. The validation accuracy increased as well and followed a pattern similar to the training accuracy, yielding final values close to or equal to training accuracy. The close association between the training accuracy and validation accuracy is encouraging and indicates that the model generalizes to unseen examples without experiencing underfitting or overfitting. This also suggests that the model's learning process was effective and carried out in an orderly manner

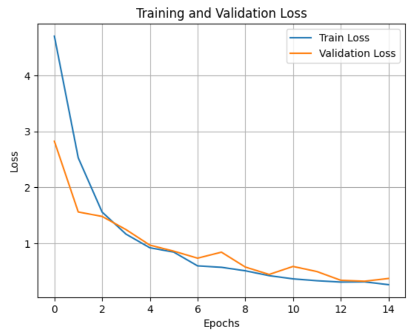


Figure III.7: Training and Validation Loss of the VGG-16 Model

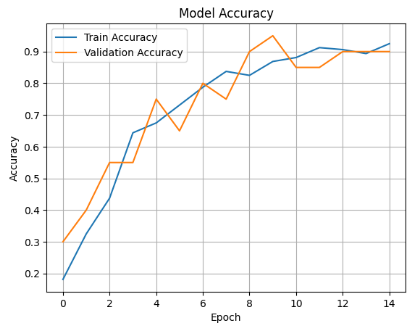


Figure III.8: Training and Validation Accuracy of the VGG-16 Model

The loss curves correspondingly showed a highly sudden decline in the training and validation loss during the first epochs, suggesting the model immediately captured valuable features. After the initial decline, the training loss and validation loss also declined overall over the course of training and had a very small distance; the distance between the training and validation loss curve is a natural behavior that could suggest that the model is not enduring overfitting with the training data. The loss curves, in effect, demonstrate the model's ability to decrease error consistently and stably. Overall, the evidence taken from the accuracy curves and loss curves demonstrate that the VGG16 model achieved consistent and stable learning with a solid feature representation, with high performance and without indications of poor generalization.

3. **Performance Analysis of the InceptionV3 Model :** The InceptionV3 model

also performed very well in classifying authors of manuscripts, obtaining an overall classification accuracy of 95%. This high overall result is indicative of the model's abilities to extract distinguishing features of differences in authors' handwriting styles. Furthermore, the high precision (0.9667), recall (0.95), and F1-score (0.9467) also show that the model performed consistently across all classes, and there is not much bias. This is further evidenced by the model's generalization and accuracy for learning features.

- **Classification Report :** The classification report displayed excellent results across most classes. For the most part, the writers obtained perfect (1.00) results for all metrics (precision, recall, and f1-score) .However there were a few classes that presented slightly less results: For the class "Ibn al-Qayyim al-Jawziyyah", the model achieved a recall of 1.0, and precision of 0.6667. This means that the model was able to establish all true samples for this specific class. However, the model also incorrectly predicted the sample of another class to belong to this class which resulted in the precision dropping due to a false positive. For the class "Al-Imam Al-Sanusi Al-Shafi'i", recall dropped to 0.50 meaning that only one of the two samples were classified appropriately and 1 sample was misclassified. Regardless, the precision was perfect (1.00), indicating that all predictions made for this class were indeed correct.

  These subtle differences would indicate that a single misprediction, has the ability to have an affect on precision as well as recall. That being said, the overall classification quality can still be deemed high and reliable.

Table III.7: Classification Raport of the InceptionV3 Model

Class	Precision	Recall	F1-score	Support
Abu_al_Hassan_abdelFattah_Al_Mahmoudi_Al_ladhqi	1.00	1.00	1.00	2
Al_Allamah Yusuf ibn Abdullah Al-Namari Al-Namari Al-Qurtubi	1.00	1.00	1.00	2
Al-Imam Abu Bakr Al-Turtushi	1.00	1.00	1.00	2
Al-Imam Al-Nawawi	1.00	1.00	1.00	2
Al-Imam Al-Sanusi Al-Shafi_i	1.00	0.50	0.6667	2
Al-Imam Ibn al-Qayyim al-Jawziyyah	0.6667	1.00	0.80	2
Ali_ibn_Mohammad_ibn_Abd_al-Samad_al-Sakhawi_al-Dimashqi	1.00	1.00	1.00	2
Hocin_bacha_siri_najl_almrfor_laho_ibrahim	1.00	1.00	1.00	2
Iskandar_jamal	1.00	1.00	1.00	2
Mobaka_om_mohmmed	1.00	1.00	1.00	2

- **Confusion matrix :** The confusion matrix provides further detail on the specific errors that the InceptionV3 model has made, providing more detail relative to the classification report. For example, one sample from the Al-Imam Al-Sanusi class was misclassified as belonging to the Ibn al-Qayyim al-Jawziyyah class, affecting the performance on both classes in precision and recall as previous noted in the classification report. In regards to the rest of the classes, all were classified perfectly, which shows that the model has strong capabilities to widely distinguish between the majorities of handwriting styles with a high degree of confidence. This positive aspect is corroborated by the macro-averaged metrics since the model's overall classification quality sustained high scores.

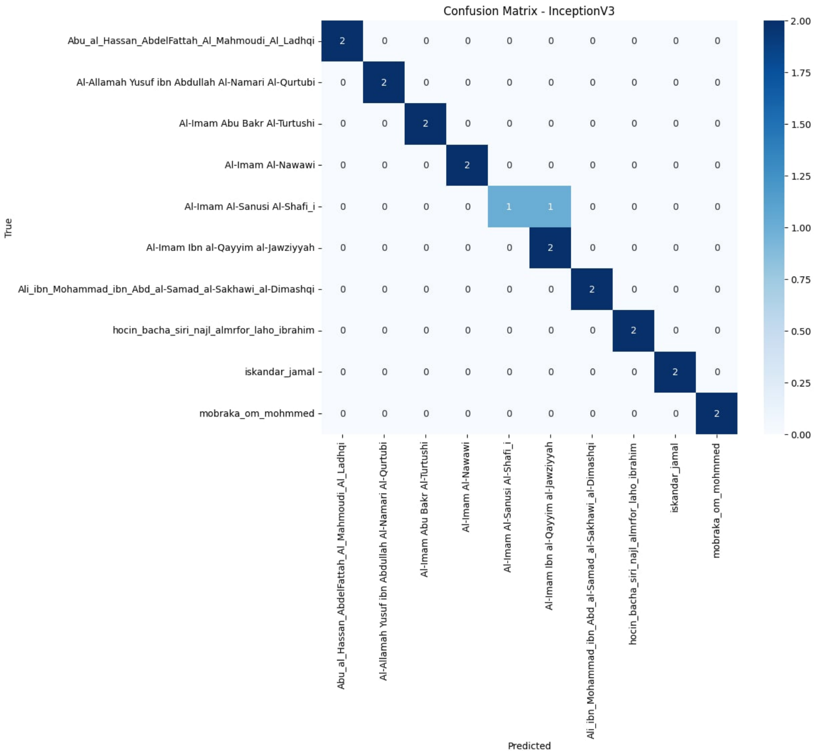


Figure III.9: Confusion matrix of the InceptionV3 Model

- **Training and Validation Accuracy and Loss Plot :** The training accuracy began to gradually ramp up in the initial epochs, which indicated learning of the initial patterns within the data. The validation accuracy began to ramp up in the early epochs quickly, indicating the models initial ability to generalize to previously unseen data. During some points of training, there seemed to be a small fluctuation in the validation accuracy, due partially to the some of the classes visually looking similar. As training progressed, the validation accuracy began to stabilize, signaling a positive outcome indicating the model is increasing its classification capabilities while still balancing learning and generalization

  During the training, the loss curve had a smooth and clear downward slope, showing that it learned and improved its errors slowly by updating the weights regularly. Although the validation loss is more variable than training loss, overall it had a downward trend with some inconsequential peaks which is typical behavior when working with smaller datasets. Specifically, it is common for the validation loss curve to be less stable because some minor variations may occur due to small samples of data representing smaller

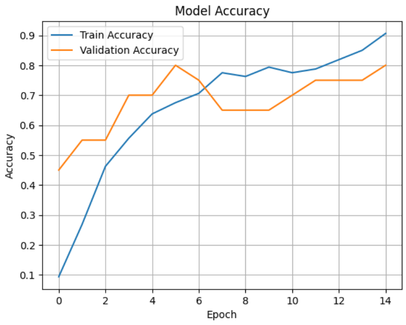


Figure III.10: Raining and Validation Accuracy of the InceptionV3 Model

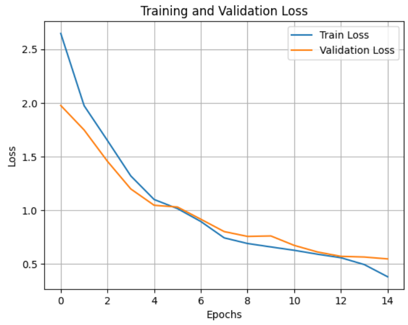


Figure III.11: Training and Validation Accuracy of the InceptionV3 Model

classes. There were no sudden spikes in the validation loss to indicate major differences in the performance of models training versus validation, meaning there were no indications of overfitting present. When both accuracy and loss curve observations were combined, the model completed effective and stable learning with a successful, strong representation of the data distinguishing features. Also, the model demonstrated clear evidence of generalization again, considering the relative small size of the dataset.

4. **Performance Analysis of the Model ResNet50 :** The test accuracy for the ResNet50 model was 75%, which is not particularly high compared to the other models and lower than VGG16 and InceptionV3, given the overall performance of each model. Although ResNet50 is theoretically supposed to have a good generalized ability, the fact that it showed strictly lower performance in this task could be due to many different factors, but ultimately, the smaller dataset may have limited the model in learning and distinguishing between classes, to the same extent as with other models. On average, for all classes, the scores for Precision, Recall, and F1-score, are as follows for the ResNet50 model. Macro Precision: 0.7000 Macro Recall: 0.7500 Macro F1-score: 0.6967 Overall, this is an average-performance and relatively balanced section overall, but does show differences in precision. Previous reflection suggests that perhaps the model was limited in learning, however, (compared to also judging patterns within single or several other classes). Thus, we must take any overall scores with appropriate caution, as there may still be examples where the model classifies incorrectly (and were of lower quality overall).

- **Classification Report** The ResNet50 model's classification report shows significant variation across classes. While several authors were classified perfectly (1.00) across all metrics, including:
  - Abu al-Hassan AbdelFattah Al-Mahmoudi Al-Ladhqi
  - Al-Allamah Yusuf ibn Abdullah Al-Qurtubi
  - Ali ibn Mohammad al-Sakhawi al-Dimashqi
  - Hocin Bacha Siri Najl al-MarfourLaho Ibrahim
  - Mobraka Om Mohmmed

  The model clearly struggled with other classes and delivered extremely poor performance For example:

  - **Al-Imam Al-Nawawi** and **Al-Imam Al-Sanusi Al-Shafi'i** had precision, recall, and F1-scores of **0.00**, which indicates that not a single sample from these classes was correctly classified.
  - **Al-Imam Abu Bakr Al-Turtushi** achieved a precision of 1.00, but a recall of 0.50, indicating that one sample was misclassified.
  - **Al-Imam Ibn al-Qayyim al-Jawziyyah** recorded a precision of 0.33 and recall of 1.00, meaning that all true samples were identified, but the model also misclassified samples from other classes into this category.
  - **Iskandar Jamal** had a precision of 0.67 and a recall of 1.00, reflecting a similar misclassification pattern. This variability in performance across classes suggests there is imbalance, which may be caused by:
  - **V**isual similarity in certain handwriting styles.
  - **U**neven or limited number of samples per class.
    These factors challenge the model's ability to accurately differentiate between classes. Even though the model performed very well for certain classes, its failure to correctly identify others significantly reduces the overall reliability of the classification results. This emphasizes the necessary to:
    This emphasizes the need to:
    * Refine or retrain the model.
    * Increase the size and diversity of the dataset.

Table III.8: Classification Raport of the ResNet50 Model

Class	Precision	Recall	F1-score	Support
Abu_al_Hassan_abdelFattah_Al_Mahmoudi_Al_ladhqi	1.00	1.00	1.00	2
Al_Allamah Yusuf ibn Abdullah Al-Namari Al-Namari Al-Qurtubi	1.00	1.00	1.00	2
Al-Imam Abu Bakr Al-Turtushi	1.00	0.50	0.6667	2
Al-Imam Al-Nawawi	0.00	0.00	0.00	2
Al-Imam Al-Sanusi Al-Shafi_i	0.00	0.00	0.00	2
Al-Imam Ibn al-Qayyim al-Jawziyyah	0.3333	1.00	0.50	2
Ali_ibn_Mohammad_ibn_Abd_al-Samad_al-Sakhawi_al-Dimashqi	1.00	1.00	1.00	2
Hocin_bacha_siri_najl_almrfor_laho_ibrahim	1.00	1.00	1.00	2
Iskandar_jamal	0.6667	1.00	0.80	2
Mobaka_om_mohmmed	1.00	1.00	1.00	2

In doing so, a more balanced and reliable classification performance across all classes can be achieved.

- **Confusion Matrix** The confusion matrix for the ResNet50 model shows even more details about its performance than the classification report provided earlier. This model achieved perfect classification for several classes, but some misclassifications had an impact on the overall performance. For example, the model misclassified one of the samples from the Al-Imam Abu Bakr Al-Turtushi class as belonging to the Al-Imam Ibn al-Qayyim al-Jawziyyah class, which accounted for the results in the last report. For the Al-Imam Al-Nawawi class the model classified incorrectly both samples as Al-Imam Ibn al-Qayyim al-Jawziyyah whereas both samples should have been classified as Al-Imam Al-Nawawi. This explains the zero precision and zero recall it scored for the Al-Imam Al-Nawawi class in the classification report. For the Al-Imam Al-Sanusi Al-Shafi'i class the model misclassified one sample as Al-Imam Ibn al-Qayyim al-Jawziyyah and one sample was predicted as iskandar_jamal. This misclassification clearly had an impact on the models' score for this class.

  Although these errors occurred, the model still scored perfect classification levels across six other classes indicating ResNet50 had some ability to distinguish between the different handwriting styles. However, there are still classes that need further improvements as they had some overlap or identifying characteristics that caused confusion.

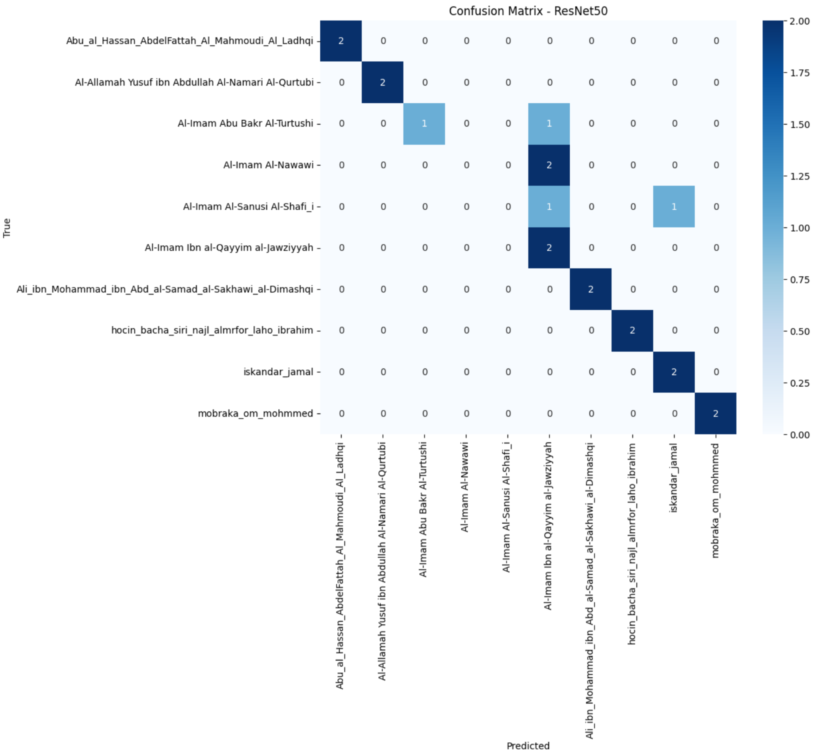


Figure III.12: Confusion Matrix of the ResNet50 Model

- **Training and Validation Accuracy and LossPlot:** The accuracy curves depict the training accuracy consistently rising until at least the latter portion of training, at which point the training accuracy is over 70%. Conversely, the validation accuracy fluctuates and remains unstable, generally falling in a range of 40% to 60% without any sign of improvement. This implies that the model has learned the training sufficiently well, but effectively generalizing to validation data, avoided overfitting training data, and improving on validation performance was a struggle for the model.

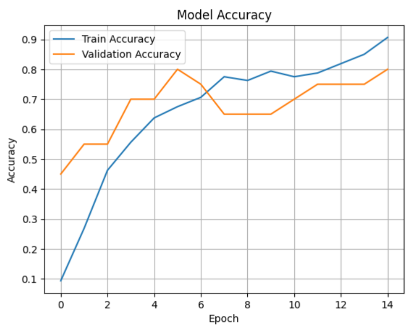


Figure III.13: Training and Validation Loss of the ResNet50 Model

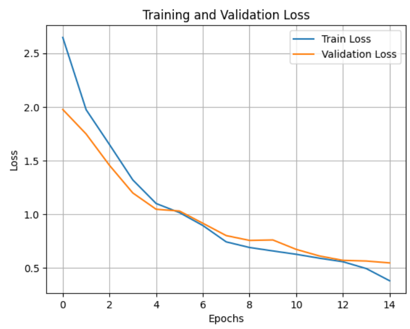


Figure III.14: Training and Validation Accuracy of the ResNet50 Model

The loss curves exhibited in the training loss exhibit a consistent downward trend over time, which is a positive indication of continued learning. The validation loss exhibited erratic behaviors and doses not exhibit a clear downward trend even indicating increases at certain points, which could potentially be another case of overfitting, where the model learned the details of the training data well, but not able to generalize. As a result of the above indicators, it can be concluded that ResNet50's performance was largely inhibited by the limited size and diversity of the data, so that the training data performed well but no effective generalization on unseen data was achieved.

III.7.3 Comparison between CNN and Transfer Learning models (Macro AVG)

The table and chart provide a comparative summary of the performance of the four models (CNN, VGG-16, InceptionV3, and ResNet50) by a range of measurement methods (accuracy, recall, precision, F1-score and loss). It is clear that VGG-16 and InceptionV3 showed the best performance across almost all measurement methods. This can be attributed to their pre-trained weights and transfer learning methods. However, despite its simplicity and small size (200 images), the CNN architecture showed strong, consistent, and stable performance across all classification tasks with an accuracy of 90%. This is reflective of the CNN's ability to pull distinctive visual features from the manuscripts. ResNet50 was the weakest model, showing that further tuning of parameters, or more data was needed for better model performance.

Comparison Table

Table III.9: Comparative Evaluation of CNN, VGG16, InceptionV3, and ResNet Based on Performance Metrics

	Accuracy	**Precision**	**Recall**	**F1-Score**	**Loss**
CNN	0.9000	0.8500	0.9000	0.8667	0.5346
VGG16	0.9500	0.9600	0.9500	0.9400	0.3147
InceptionV3	0.9500	0.9600	0.9500	0.9400	0.2189
ResNet50	0.7500	0.7000	0.7500	0.6900	0.7600

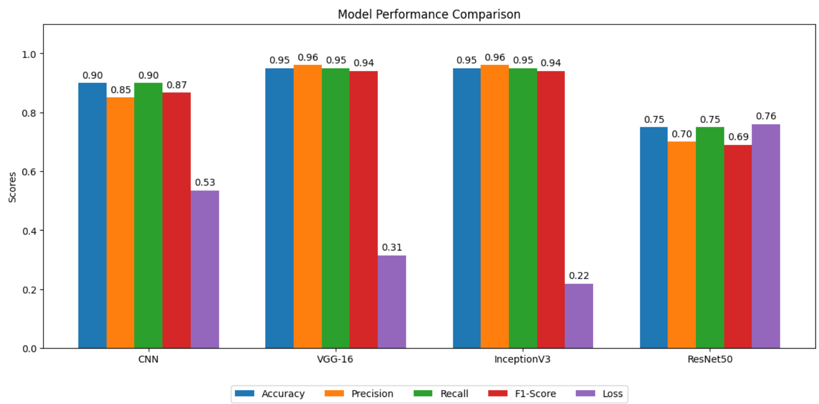


Figure III.15: Performance Comparison of CNN, VGG16, InceptionV3, and ResNet Models

III.7.4 Discussion

The results obtained in this study provide several insights into the challenges and opportunities in applying deep learning techniques to the authorship attribution of ancient Arabic manuscripts. One of the most significant findings is that a custom-built CNN model, despite being relatively shallow and trained on a limited dataset of only 200 images, was still capable of achieving a competitive accuracy of 90%. This indicates that with careful preprocessing, appropriate normalization techniques, and sufficient training iterations, even basic architectures can effectively learn distinctive handwriting patterns that reflect individual authorship traits. However, the study also revealed the clear benefits of transfer learning, particularly when using well-established architectures such as VGG16 and InceptionV3. Both models reached an accuracy of 95%, outperforming the custom CNN by a margin of 5%. This performance gain can be attributed to the pre-initialization of model weights on the large-scale ImageNet dataset, which enables the models to leverage previously learned features relevant to visual recognition tasks. Transfer learning proved especially beneficial in our case due to the limited size and diversity of the dataset, highlighting its value in scenarios where data collection is constrained. Further evidence of the advantage of transfer learning is found in the confusion matrices, where pre-trained models exhibited clearer diagonal patterns, suggesting a higher consistency in correctly identifying authors. This implies that the pre-trained models were better able to generalize and distinguish subtle differences in handwriting style—differences that are often nuanced in Arabic script due to calligraphic variations and complex ligatures. Nevertheless, the overall performance of all models was impacted by several limitations, most notably the small size of the dataset and the challenges in sourcing and labeling Arabic manuscript images. Unlike Latin-based manuscripts, which are extensively documented and well-annotated in existing datasets, Arabic manuscripts suffer from a lack of structured, publicly available corpora. The process of data collection was manual and labor-intensive, with inconsistencies in manuscript quality, ink tone, and writing tools adding further complexity to the task. The data augmentation techniques applied during preprocessing played a crucial role in mitigating overfitting and enhancing the model's generalization capabilities. Image resizing, normalization, mean subtraction, and standardization were essential in harmonizing visual features across varying image qualities. Moreover, the augmentation strategies simulated natural variability in handwriting, which is particularly important for modeling the stylistic diversity found in historical Arabic documents. In conclusion, this discussion reaffirms the importance of both model design and data preparation. While deep learning offers powerful tools for authorship attribution, the success of such models depends heavily on the availability of clean, diverse, and well-labeled datasets, as well as on the adoption of effective training and preprocessing strategies

Conclusion

Throughout this chapter, we detailed the comprehensive process undertaken to develop an automatic classification system for Arabic manuscript images using deep learning methodologies. Preprocessing techniques and data augmentation were crucial in enhancing the quality and diversity of the dataset, thereby enabling the models to learn more effectively from limited data. The application of pre-trained convolutional neural networks significantly improved classification performance compared to the baseline custom model. These promising results underscore the viability of CNN-based approaches in the context of Arabic handwriting analysis. Moreover, this work represents an important step toward the broader objective of automating the identification of manuscript authors, with potential applications in the digital preservation, cataloguing, and scholarly study of historical handwritten documents.

Conclusion General

This study set out to investigate the potential of deep learning techniques for the automatic identification of authors of ancient Arabic manuscripts—an endeavor of both scholarly and cultural importance. Authorship attribution in historical Arabic documents plays a vital role in preserving the intellectual heritage of the Arab world, aiding digital archiving efforts, and facilitating academic research in fields such as palaeography and manuscript studies. Due to the absence of publicly available datasets tailored to this task, we constructed a custom dataset comprising 10 authors, each represented by 20 manuscript images. Although modest in size, this dataset served as a foundational benchmark for evaluating the feasibility of handwriting-based authorship identification using deep learning. A variety of preprocessing steps—including image resizing, pixel normalization, and mean subtraction—were applied to improve input quality, while data augmentation techniques addressed the challenge of data scarcity and promoted better model generalization. These procedures enabled the models to generalize better despite the inherent limitations in sample size. Both a custom-designed Convolutional Neural Network (CNN) and several pre-trained deep learning models were implemented and evaluated. . The results highlight the value of deep learning in the domain of Arabic manuscript analysis. Despite the limitations imposed by the small dataset and variability in manuscript quality, the models were able to generalize effectively and distinguish between the writing styles of different authors. This confirms that Ara handwriting carries sufficient stylistic uniqueness to enable authorship classification, even in challenging conditions.

Future Work and Perspectives

Building on these promising findings, several directions are recommended for future work:

- **Dataset Expansion:** Increase the number of authors and images, incorporating manuscripts from diverse historical periods and regions to improve model robustness.
- **Advanced Architectures:** Explore hybrid models that combine deep learning with stylometric and linguistic analysis.
- **Multimodal Approaches:** Integrate contextual metadata such as date, geographic origin, or type of writing tool to enrich the classification process.
- **Image Enhancement:** Utilize advanced restoration tools to improve the legibility of degraded manuscripts.
- **Real-World Applications:** Apply the model to anonymous or disputed manuscripts to assess its capacity for intelligent authorship suggestion.
- **User Interface Development:** Create an interactive platform for researchers to upload manuscripts and benefit from automated analysis tools.

In conclusion, this research demonstrates the potential of deep learning as a powerful tool for automatic authorship attribution in ancient Arabic manuscripts. It establishes a solid foundation for further exploration and innovation in the intersection of artificial intelligence, digital humanities, and cultural heritage preservation.

Bibliography

[1] Arab Papers Foundation, ”Definition of Manuscripts and Their Types.” [Online]. Available: https://aawraq.com/article/e [Accessed: Apr. 12, 2025].

[2] Najah.net, ”Arabic manuscripts: Their history and types.” [Online]. Available: https://www.najah.edu [Accessed: Feb. 20, 2025].

[3] A. S. Osborn, *Questioned Documents*, 2nd ed. Boyd Printing Company, 1929. [Online]. Available: https://archive.org/details/documequestioned00osborich [Accessed: Feb. 3, 2025].

[4] M. Saini, ”Variability in Handwriting Patterns among Ethnic Groups,” Int. J. Humanit. Soc. Sci. (IJHSS), vol. 2, no. 12, pp. 71–76, 2013

[5] N. Arica and F. T. Yarman-Vural, ”An overview of character recognition focused on off-line handwriting,” IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 31, no. 2, pp. 216–233, 2001.

[6] M. Koppel, J. Schler, and S. Argamon, ”Computational methods in authorship attribution,” J. Am. Soc. Inf. Sci. Technol., vol. 60, no. 1, pp. 9–26, 2009.

[7] E. Stamatatos, ”A survey of modern authorship attribution methods,” J. Am. Soc. Inf. Sci. Technol., vol. 60, no. 3, pp. 538–556, 2009.

[8] P. Juola, ”Authorship Attribution,” Found. Trends Inf. Retr., vol. 1, no. 3, pp. 233–334, 2006.

[9] U.-V. Marti and H. Bunke, ”The IAM-database: An English sentence database for offline handwriting recognition,” Int. J. Doc. Anal. Recognit., vol. 5, no. 1, pp. 39–46, 2002.

[10] M. Bulacu and L. Schomaker, ”Text-independent writer identification and verification using textural and allographic features,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 4, pp. 701–717, 2007.

[11] S. Fiel and R. Sablatnig, ”Writer identification and retrieval using a con-

volutional neural network," in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW), pp. 1010–1016, 2015.

[12] A. Fischer, A. Keller, V. Frinken, and H. Bunke, "Lexicon-free handwritten word spotting using character HMMs," Pattern Recognit. Lett., vol. 33, no. 7, pp. 934–942, 2011.

[13] F. Kleber, G. A. Fink, and R. Ingold, "The CVL database: An off-line database for writer retrieval, writer identification and word spotting," in Proc. 12th Int. Conf. Doc. Anal. Recognit. (ICDAR), pp. 560–564, 2013.

[14] S. A. Mahmoud, A. Alghamdi, W. G. Al-Khatib, and A. Al-Harbi, "KHATT: Arabic offline handwritten text database," in Proc. 14th Int. Conf. Frontiers Handwriting Recognit., pp. 186–191, 2014.

[15] S. Al-Maadeed, A. Hassaine, and S. Al-Maadeed, "Writer identification using offline Arabic handwriting," Pattern Recognit. Lett., vol. 35, pp. 68–77, 2014.

[16] L. Schomaker and M. Bulacu, "Automatic writer identification using connected-component contours and edge-based features of uppercase western script," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 6, pp. 787–798, 2004.

[17] M. Koppel, J. Schler, and S. Argamon, "Computational methods in authorship attribution," J. Am. Soc. Inf. Sci. Technol., vol. 60, no. 1, pp. 9–26, 2009.

[18] F. A. Batarseh, "Artificial Intelligence," in Encyclopedia of Big Data, pp. 1–5, 2022.

[19] A. Althbiti and X. Ma, "Machine Learning," in Encyclopedia of Big Data, pp. 1–5, 2020.

[20] G. Folino, M. Guarascio, and M. A. Haeri, "Deep Learning on Big Data," in Encyclopedia of Big Data Technologies, pp. 1–10, 2022.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Adv. Neural Inf. Process. Syst., 2012. [Online]. Available: https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b [Accessed: Apr. 12, 2025].

[22] S. S. Tekkam, M. S. Koti, and P. S. Hiremath, "Object Detection and Tracking Using Deep Learning: A Survey," Technologies, vol. 12, no. 2, Art. no.

15, 2024.

[23] A. Gurnani, V. Mavani, V. Gajjar, and Y. Khandhediya, "Flower Categorization using Deep Convolutional Neural Networks," arXiv:1708.03763, 2017. [Online]. Available: https://arxiv.org/abs/1708.03763 [Accessed: May. 3, 2025].

[24] S. Ahlawat, A. Choudhary, A. Nayyar, S. Singh, and B. Yoon, "Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN)," Sensors, vol. 20, no. 12, Art. no. 3344, 2020.

[25] M. M. Khayyat and L. A. Elrefaei, "Towards author recognition of ancient Arabic manuscripts using deep learning: A transfer learning approach," Int. J. Comput. Digit. Syst., vol. 90, no. 5, 2020.

[26] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, "Conceptual understanding of convolutional neural network - A deep learning approach," Procedia Comput. Sci., vol. 132, pp. 2018–2025, 2018.

[27] S. Albelwi and A. Mahmood, "A Framework for Designing the Architectures of Deep Convolutional Neural Networks," Entropy, vol. 19, no. 6, Art. no. 6, 2017.

[28] A. Ossama, "Digital Image Course." [Unpublished course material].

[29] A. H. Reynolds, "Anh H. Reynolds." [Online]. Available: https://anhreynolds.com [Accessed: Feb. 20, 2025].

[30] M. Purwono, "Understanding of Convolutional Neural Network (CNN): A Review," Int. J. Robot. Control Syst., vol. 2, no. 4, pp. 739–748, 2022.

[31] S. Sakib, N. Ahmed, A. J. Kabir, and H. Ahmed, "An overview of convolutional neural network: Its architecture and applications," Preprints, 2019. [Online]. Available: https://doi.org/10.20944/preprints201811.0546.v4 [Accessed: Apr. 12, 2025].

[32] J. Li, Y. Hai, and S. Yin, "A survey of artificial intelligence for industrial detection," Ann. Data Sci., vol. 12, no. 2, pp. 799–827, 2024.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 770–778, 2016.

[34] S. Albahli and T. Nazir, "AI-CenterNet CXR: An artificial intelligence (AI) enabled system for localization and classification of chest X-ray disease," Front. Med., vol. 9, Art. no. 955765, 2022.

[35] C. Szegedy et al., ”Going deeper with convolutions,” arXiv:1409.4842, 2014. [Online]. Available: https://arxiv.org/pdf/1409.4842 [Accessed: Apr. 3, 2025].

[36] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[37] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.

[38] Alreq AlManshour, ”Alreq AlManshour: Electronic Catalog of Middle Eastern Manuscripts,” [Online]. Available: https://app.alreq.com/ar [Accessed: May. 12, 2025].

[39] Alkitabdar, ”Islamic Manuscripts Directory,” [Online]. Available: https://alkitabdar.com/manuscripts/ [Accessed: May. 12, 2025].

[40] ILM Foundation for Heritage Revival and Digital Services, ”ILM Arabia Platform for Arabic-Islamic Manuscript Studies,” 2005. [Online]. Available: https://ilmarabia.com/home [Accessed: March. 14, 2025].