International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Outlier Detection in Signatures Using Deep Learning

Omjee Tiwary[1], Shailesh Maurya[2], Vivek Prajapati[3]

[1,2,3]Student, Computer Science Engineering with specialization Artificial Intelligence and Machine Learning, Greater Noida Institute Of Technology

Abstract

This research paper presents a deep learning-based approach for detecting outliers in handwritten signatures, which plays a crucial role in signature verification systems. Signature forgery detection is a challenging task due to variations in genuine signatures caused by writing style, pressure, and environmental factors. The proposed system utilizes an autoencoder-based deep learning model to learn the normal patterns of genuine signatures and identify anomalies (outliers) based on reconstruction error. The methodology involves preprocessing signature images, feature extraction using convolutional neural networks (CNNs), and training an autoencoder to reconstruct genuine signatures. During testing, signatures that deviate significantly from learned patterns are classified as outliers. A statistical threshold, calculated using mean and standard deviation of reconstruction errors, is applied to distinguish between genuine and forged signatures.

Keywords: Outlier Detection, Signature Verification, Deep Learning, Autoencoder, CNN, Anomaly Detection.

1. Introduction

- Outlier detection in handwritten signatures has emerged as a significant research problem in the field of pattern recognition and biometric authentication. Signatures are widely used for identity verification in banking, legal documentation, and administrative processes. However, the presence of forged or anomalous signatures poses a serious threat to the reliability of such systems. It is observed that genuine signatures often exhibit natural variations due to factors such as writing speed, pressure, and emotional state, making the task of distinguishing between authentic and forged signatures more challenging.
- Traditional signature verification methods rely heavily on handcrafted features and statistical techniques. These approaches may fail to capture complex patterns present in signature data, especially when dealing with skilled forgeries. Therefore, the use of deep learning techniques has gained attention due to their ability to automatically learn discriminative features from raw data. In particular, Convolutional Neural Networks and autoencoder architectures have shown promising results in modelling the intrinsic characteristics of genuine signatures.
- In this research, a deep learning-based approach is proposed for detecting outliers in signature images. The system is designed to learn the normal patterns of genuine signatures and identify deviations that indicate anomalies or forgeries. It is observed that autoencoders can effectively reconstruct genuine signatures with minimal error, while forged signatures result in higher reconstruction errors. A

statistical threshold based on mean and standard deviation is used to classify signatures as genuine or outliers.

- The proposed methodology includes preprocessing of signature images, feature extraction using deep neural networks, and anomaly detection using reconstruction error analysis. This approach reduces dependency on manual feature engineering and improves the robustness of signature verification systems. The study aims to enhance accuracy and reliability in real-world applications where secure identity verification is essential.

2. Abbreviations and Acronyms

In this research paper, several abbreviations and acronyms are used to represent commonly referenced terms in the domain of deep learning and signature verification. Each abbreviation is defined at its first occurrence in the text to ensure clarity and understanding.

DL (Deep Learning): A subset of machine learning that uses neural networks with multiple layers to learn patterns from data.

CNN (Convolutional Neural Network): A class of deep neural networks commonly used for image processing tasks such as feature extraction from signature images.

AE (Autoencoder): A type of neural network used for unsupervised learning, which encodes input data into a compressed representation and then reconstructs it. It is used in this research for anomaly detection.

OD (Outlier Detection): The process of identifying data points that significantly deviate from normal patterns.

SV (Signature Verification): The task of verifying whether a given signature is genuine or forged.

ML (Machine Learning): A broader field of artificial intelligence that enables systems to learn and improve from experience without being explicitly programmed.

MSE (Mean Squared Error): A loss function used to measure the difference between original and reconstructed images in the autoencoder model.

SD (Standard Deviation): A statistical measure used along with mean to determine the threshold for identifying outliers.

ROI (Region of Interest): The specific portion of a signature image selected for processing and analysis.

All abbreviations are consistently used throughout the paper after their initial definition to maintain readability and avoid ambiguity.

3. Units

The **SI (International System of Units)** is used as the primary system throughout the paper, as it is widely accepted in scientific research. Since this study mainly deals with digital images and deep learning models, most values are related to image size, computation time, and system performance rather than physical measurements. For example, image dimensions are usually represented in **pixels**, while any physical measurements (if required) are expressed in units like **centimetres (cm)**. Time-related values such as training or processing time are written in **seconds (s)**.

Care is taken to avoid mixing SI and CGS units, because combining different unit systems can make equations confusing and difficult to interpret. In cases where different types of units are unavoidable, each unit is clearly mentioned so that there is no ambiguity.

Also, consistency is maintained while writing units. Either the full name or the standard abbreviation is used, but both are not mixed together in the same expression. For example, terms like “pixels per centime-

tre" or "Px/cm" are used properly instead of combining formats incorrectly.
All units are written with a proper space between the number and the unit, such as **"256 Px"**, **"10 cm"**, and **"0.5 s"**. Correct capitalization is also followed while writing units to maintain standard formatting.

4. Equations

- Whenever an equation involves division, fractions, or multiple terms, the built-in **equation editor** of the word processor is used. This helps in maintaining proper formatting and avoids confusion while reading complex expressions. All equations are **left aligned** to keep the document structure clean and uniform.
- It is preferred to assign **serial numbers** to equations for easy reference within the text. These numbers are written in parentheses and are usually placed towards the right side of the page. When multiple equations are included, their numbering is aligned at the same position to maintain consistency.
- No italic styling is applied to equations, and the **font size remains the same as the normal text** of the document. A blank line is added before and after each equation to clearly separate it from the surrounding content.
- For simple equations that are not created using the equation editor, standard symbols are used. The multiplication sign is written as **"×"** instead of "*", and division is written as **"÷"** instead of "/". Also, decimal values are written properly with a leading zero, such as **"0.25"** instead of ".25".
- Some example equations relevant to this research are given below:

Mean of reconstruction error:
$$\mu = (1 \div N) \times \Sigma\, Ei \qquad (1)$$
Standard deviation:
$$\sigma = \sqrt{[(1 \div N) \times \Sigma\, (Ei - \mu)^2]} \qquad (2)$$
Threshold for outlier detection:
$$T = \mu + 2 \times \sigma \qquad (3)$$
In these equations, **Ei** represents the reconstruction error of each signature sample, **N** is the total number of samples, **μ** is the mean error, and **σ** is the standard deviation. Any signature with an error greater than the threshold **T** is considered an outlier.

5. Headings

All headings use the same **font style and size** as the normal text (Times New Roman, 12 pt). The only formatting applied to headings is **bold**, so they stand out clearly from the rest of the content. No underline or italic styles are used, as they can make the document look cluttered.
Headings can be written either with numbering or without numbering, but it is generally better to use **numbered headings** for better organization. Only standard numbering (like 1, 1.1, 1.1.1) is used, and Roman numerals or alphabets are avoided. This makes it easier to structure sections such as Introduction, Methodology, Results, and Conclusion.
For sub-sections, **hierarchical numbering** can be used. For example:
This helps in clearly showing the relationship between different sections of the paper.
To maintain proper formatting, the **"Keep with next paragraph"** option is enabled for all headings. This ensures that a heading always appears on the same page as its content, rather than being separated across pages

6. Figures and Tables

All figures and tables include proper **captions** using the caption feature available in the word processor. The captions are written in the same font style and size as normal text, without using bold, italic, or underline formatting. For better clarity, **Title Case** is used in captions, such as *"Figure 1: Autoencoder Architecture for Signature Reconstruction"*.

Both figures and tables are **centre aligned**, along with their captions, to maintain a neat and professional layout. It is recommended to assign **numbers** to all figures and tables so they can be easily referred to within the text. Instead of short forms like "Fig. 1", full terms like **"Figure 1"** and **"Table 1"** are used.

Captions are preferably placed **above the figures and tables**, making it easier for readers to understand the content before viewing it. Also, text is not written alongside figures or tables; each element is placed on a separate line without wrapping.

Tables are designed to be simple and readable. No background colours are applied to cells, rows, or columns. The size of table cells is kept just enough to fit the content, ensuring that tables remain compact. Column headings and important rows like totals are written in **bold** for better visibility.

The font size inside tables is kept the same as normal text. However, if a table is too wide, the font size may be reduced slightly (for example, to 10 pt). If it still does not fit properly, the table can be divided into smaller parts.

Images used in figures, such as signature samples or model diagrams, are maintained in their **original proportions**. They are not stretched or compressed, ensuring clarity and proper resolution. If multiple small images (like signature samples) are used, they can be arranged in a single row using a table layout.

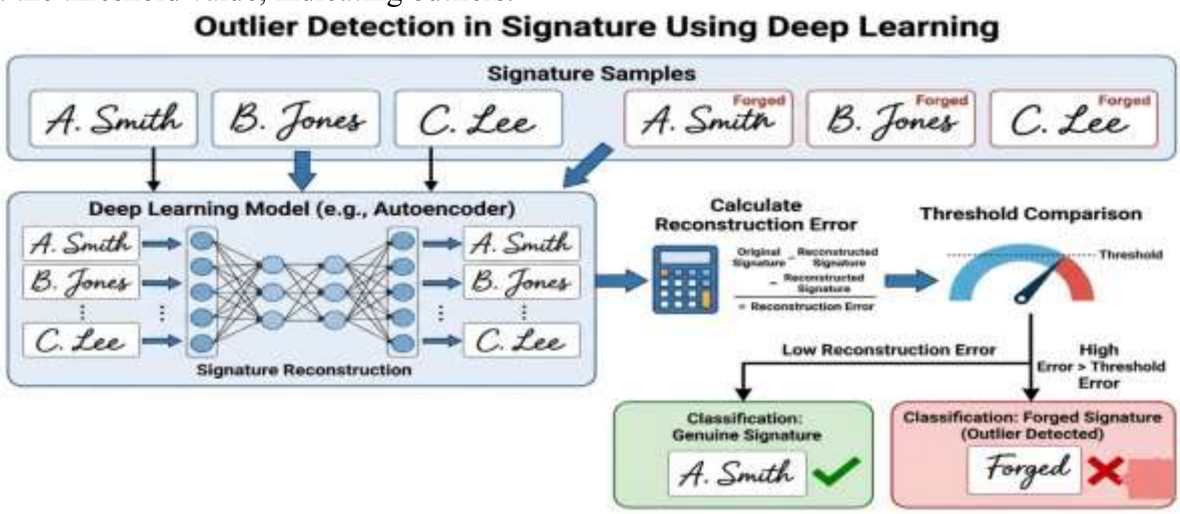
Blank lines are added before and after each figure and table to clearly separate them from the surrounding text.

An example table relevant to this research is shown below:

Table 1: Reconstruction Error Values for Signature Samples

Sample Type	**Sample 1**	**Sample 2**	**Sample 3**
Genuine	0.12	0.15	0.10
Forged	0.45	0.52	0.48
Threshold	0.30	0.30	0.30

The above table shows that genuine signatures have lower reconstruction error, while forged signatures exceed the threshold value, indicating outliers.



7. Some Common Mistakes

While preparing this research paper on *Outlier Detection in Signature Using Deep Learning*, care has been taken to avoid common writing and formatting mistakes that can affect clarity and professionalism.

One common mistake is the incorrect use of symbols. For example, using "0" (zero) or "O" with superscript formatting instead of the proper **degree symbol (°)**. Although this research does not heavily involve temperature or angular measurements, whenever such values are mentioned, the correct symbol is used to maintain standard formatting.

Another frequent error is related to the Latin abbreviation **"et al."**, which is often used in citations. It is important to note that there is **no full stop after "et"**, and the correct format is *"et al."*, not "et. al.". This rule is followed consistently throughout the paper while referring to multiple authors.

There is also careful attention to the correct usage of **"i.e."** and **"e.g."**, as they are often confused. The abbreviation **"i.e."** means *"that is"* and is used to clarify or restate something, while **"e.g."** means *"for example"* and is used to provide examples. These terms are used appropriately in the context of explaining concepts like model techniques, datasets, or evaluation metrics.

Acknowledgement

The authors would like to express their sincere gratitude to their project guide, **Mr. Amrish Dubey**, for his valuable guidance, continuous support, and encouragement throughout the development of this research work on *Outlier Detection in Signature Using Deep Learning*. His insights and suggestions greatly contributed to the successful completion of this project.

The authors, **Omjee Tiwary**, **Shailesh Maurya**, and **Vivek Prajapati**, are also thankful to their institution for providing the necessary resources, facilities, and a supportive environment to carry out this research.

References

References within Main Content of the Research Paper

While referring to other research papers or books in the main content, **citation numbers are written in square brackets**, such as **[1]**. These numbers are placed **before the full stop** at the end of the sentence.

Where needed, the **author's name and year of publication** can also be mentioned in brackets, for example: *(John Smith, 2020)*. If a paper has multiple authors, only one author's name is written followed by **"et al."**, such as *(Smith et al., 2020)*.

If more than one reference is used at the same place, they can be written together like **[1, 2]**, separated by a comma and space. While referring to references, only the number is used, such as **[3]**, instead of writing "Ref. [3]" or "reference [3]".

Also, references are not used as the subject of a sentence. For example, instead of writing *"as explained in [1]"*, it is better to write *"as Smith (2020) explains"*.

References in the Reference List

At the end of the research paper, a complete list of references is provided. This is usually written as an **end-note section**, which is preferred over footnotes when references are used multiple times.

References are numbered using a simple format like **1., 2., 3.**, and each reference includes details such as **authors' names, title of the paper, journal or publisher name, year of publication, volume, issue, and page numbers**.

The following points are followed while writing references:

1. No part of the reference is written in italic style.

2. Each part of the reference is separated by a comma.
3. A full stop is added at the end of each reference (except after URLs).
4. If a URL is included, it should be complete and directly link to the exact paper or resource.
5. Titles are written in proper format (Title Case or Sentence case), not fully uppercase or lowercase.
6. Author names are written consistently in one format throughout the list.