Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 10, October : 2024

ENHANCED MECHANISM OF SIGNATURE RECOGNITION AND AUTHENTICATION USING MACHINE LEARNING ALGORITHMS

[1]K. Jamuna, [2]A. Mounika, [3]G. Dileep, [4]G. Jyothi

[1,2]Assistant Professor, Department of CSE(AI&ML), Vignan's Institute of Management and Technology For Women , Ghatkesar, Medchal, Telangana

[3]Assistant Professor, Department of CSE, Vignan's Institute of Management and Technology For Women , Ghatkesar, Medchal, Telangana

[4]Assistant Professor, Department of CSE, Nalla Narasimha reddy group of institutions, Narapally, Chowdaryguda, Telangana

E-Mail: [1]jamuna.kongari1@gmail.com, [2]andemounikaab@gmail.com, [3]dilleepgangavath.42@gmail.com, [4]gugulothu.jyothi@gmail.com

ABSTRACT

Signature recognition and authentication play a critical role in identity verification for a wide range of applications, including banking, legal documentation, and access control. Traditional methods of signature verification rely on human experts or basic rule-based systems, which are often prone to errors, fraud, and inconsistency. To address these challenges, the implementation of machine learning algorithms offers an enhanced mechanism for accurate and efficient signature recognition and authentication. This paper proposes an advanced approach to signature recognition and authentication using machine learning algorithms. The system leverages various techniques such as support vector machines (SVM), convolutional neural networks (CNNs), and recurrent neural networks (RNNs) to analyze both static and dynamic signature features, including stroke patterns, pressure, and speed. By training models on large datasets of genuine and forged signatures, the system learns to distinguish between authentic and fraudulent signatures with high accuracy. Experimental results demonstrate significant improvements in signature recognition accuracy, reducing false positives and false negatives compared to traditional verification methods. The use of machine learning enhances the robustness of the system by adapting to variations in signature styles, improving security, and minimizing the risk of forgery. This method provides a scalable and automated solution for signature-based authentication, applicable across various industries that require reliable identity verification

Keywords: *Signature Recognition*, **:** *Signature Authentication, Machine Learning, SVM, CNN, forgery detection, identity verification, RNN.*

INTRODUCTION

Machine learning is the study of computer algorithms that improve automatically through experience and using data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks. features like textures and shapes, a CNN takes just the image's raw pixel data as input and "learns" how to extract these features, and ultimately infer what object they constitute [1]. Machine learning involves computers discovering how they can perform tasks without being explicitly programmed to do so. It involves computers learning from data provided so that they carry out certain tasks. For simple tasks assigned to computers, it is possible to program algorithms telling the machine how to execute all steps required to solve the problem at hand; on the computer's part, no learning is needed. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine develop its own algorithm, rather than having human programmers specify every needed step. In the existing system, they need only one pre-given signature to verify whether the given input signature is a true one or not [2]. Here comes the first disadvantage where if they use a single signature let's assume that the user may be in a hurry and he signed it in a hurry sometimes that lead to improper or imperfect signatures so in this condition this system fails to compare them perfectly whereas in our system it takes 5 to 6 signatures let us assume the same condition that a person is in hurry in that situation he might sign one or two improperly but any one of them will be correct so it compares the new signature with duplicate one's if it matches to any one of them it will be considered as correct one or else wrong one. So to get the proper outcome we can modify the approach of comparing the signatures in a more proper way. Here a proposed system is provided where the normalization of the signature image is done and system checks whether the signature matches with original signature.Test signature is recognized with the given input training set using CNN . Then forgery detection algorithms (Softmax regression model) and herris algorithm are enforced on this classified image wherein corners points of the signatures are validated and also to ensure whether signature boundaries are perfectly aligned or not , based on the values System will display the result accordingly with an accuracy of 60%-80% depending on image quality, image lighting and image background [3].

LITERATURE SURVEY

Signature recognition and authentication have long been critical components in verifying identity and authorizing transactions across various sectors, including finance, legal, and governmental domains. Traditional methods of signature verification, which often rely on manual inspection or rudimentary

feature extraction techniques, face significant limitations such as susceptibility to human error, forgery, and high false-positive/false-negative rates [4]. This literature review explores the advancements in signature recognition and authentication through the application of machine learning algorithms, focusing on their ability to improve accuracy, security, and efficiency [5]. Historically, signature recognition has been categorized into two types: static (offline) and dynamic (online). Static recognition focuses on the analysis of a signature's visual appearance from scanned documents, while dynamic recognition captures temporal features such as stroke order, pressure, and velocity during the signing process. Early approaches relied heavily on rule-based methods, geometric features, and handcrafted descriptors, which were limited by their inability to adapt to intra-class variability (i.e., natural differences in an individual's signatures over time) and inter-class similarity (i.e., similarity between different people's signatures) function of Machine Learning Algorithms [6]. The emergence of machine learning provided a paradigm shift in signature recognition. Machine learning algorithms are designed to automatically learn features from data, making them far more adaptable to variations in signature characteristics [7]. Studies by Sabourin et al. (1997) demonstrated how SVMs, trained on signature features such as line orientation and curvature, significantly improved recognition rates for static signatures. Hafemann et al. (2017) applied CNNs to offline signature recognition, achieving state-of-the-art results by leveraging deep feature extraction and minimizing the need for manual feature engineering. Research by Impetigo and Pirlo (2019) explored the use of LSTMs for dynamic signature verification, demonstrating their effectiveness in capturing temporal dependencies such as pressure and speed variations, which are critical in differentiating between authentic and forged signatures. Machine learning algorithms, particularly deep learning models such as CNNs and LSTMs, have revolutionized the field of signature recognition and authentication [8]. These methods provide superior performance in detecting forgeries, adapting to variations in signatures, and reducing error rates compared to traditional methods. As research continues, these models will likely play an increasingly important role in enhancing the security and reliability of signature-based authentication systems in various industries.

METHODOLOGY

Classification in Machine Learning:

- Supervised Machine Learning
- Unsupervised Machine Learning

Supervised Machine Learning: The majority of practical machine learning uses supervised learning. Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output $Y = f(X)$.

The goal is to approximate the mapping function so well that when you have new input data (x) that you

can predict the output variables (Y) for that data.

The techniques of Supervised Machine Learning algorithms include linear and logistic regression, multiclass classification, Decision Trees and support vector machines. Supervised learning requires that the data used to train the algorithm is already labelled with correct answers. For example, a classification algorithm will learn to identify animals after being trained on a dataset of images that are properly labelled with the species of the animal and some identifying characteristics. Supervised learning problems can be further grouped into Regression and Classification problems. Both problems have as goal the construction of a succinct model that can predict the value of the dependent attribute from the attribute variables. The difference between the two tasks is the fact that the dependent attribute is numerical for regression and categorical for classification.

Regression analysis consists of a set of machine learning methods that allow us to predict a continuous outcome variable (y) based on the value of one or multiple predictor variables (x). Briefly, the goal of regression model is to build a mathematical equation that defines y as a function of the x variables. Next, this equation can be used to predict the outcome (y) on the basis of new values of the predictor variables (x). There are two types of regression models, they are

1. Single Regression model
- Linear Regression model
- Non-linear Regression model
2. Multiple Regression model
- Linear Regression model
- Non-linear Regression model

Linear regression the most simple and popular technique for predicting a continuous variable. It assumes a linear relationship between the outcome and the predictor variables.

The linear regression equation can be written as $y = b0 + b*x + e$, where:

- b0 is the intercept,
- b is the regression weight or coefficient associated with the predictor variable x.
- e is the residual error

Technically, the linear regression coefficients are determined so that the error in predicting the outcome value is minimized. This method of computing the beta coefficients is called the Ordinary Least Squares method. When you have multiple predictor variables, say x1 and x2, the regression equation can be written as $y = b0 + b1*x1 + b2*x2 + e$. In some situations, there might be an interaction effect between

some predictors, that is for example, increasing the value of a predictor variable x1 may increase the effectiveness of the predictor x2 in explaining the variation in the outcome variable.

Note also that, linear regression models can incorporate both continuous and categorical predictor variables . In some cases, the relationship between the outcome and the predictor variables is not linear. In these situations, you need to build a Non-linear Regression, such as polynomial and spline regression. When you have multiple predictors in the regression model, you might want to select the best combination of predictor variables to build an optimal predictive model. This process called model selection, consists of comparing multiple models containing different sets of predictors in order to select the best performing model that minimize the prediction error. Linear model selection approaches include best subsets regression and stepwise regression. You can apply all these different regression models on your data, compare the models and finally select the best approach that explains well your data. To do so, you need some statistical metrics to compare the performance of the different models in explaining your data and in predicting the outcome of new test data. The best model is defined as the model that has the lowest prediction error. The most popular metrics for comparing regression models, include:

Root Mean Squared Error: RMSE measures the model prediction error. It corresponds to the average difference between the observed known values of the outcome and the predicted value by the model. RMSE is computed as RMSE = mean((observeds - predicteds)^2) %>% sqrt(). The lower the RMSE, the better the model.

Lasso Regression: The "LASSO" stands for Least Absolute Shrinkage and Selection Operator. Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models with fewer parameters. This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

Unsupervised Machine Learning- Unsupervised learning is the training of a machine using information that is neither classified nor labelled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data. Unlike supervised learning, no training will be given to the machine. Therefore, the machine is restricted to find the hidden structure in unlabelled data by itself.

Unsupervised learning is classified into two categories of algorithms:

Clustering: A clustering problem is where you want to discover the inherent groupings in the data, such

as grouping customers by purchasing behaviour.

Association: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

SoftMax Regression Model:

Linear regression, logistic regression and softmax regression models can be derived from a general linear model. In a logistic regression model, the outcome or 'y' can take on binary values 0 or 1 where as in softmax regression, the outcome 'y' can take on multiple values.ITconcepts like partial differentiation, maximum likelihood function, gradient descent and matrix multiplication. .

Random Search Algorithm

A random search algorithm refers to an algorithm that uses some kind of randomness or probability (typically in the form of a pseudo-random number generator) in the definition of the method, and in the literature, may be called a Monte Carlo method or a stochastic algorithm. The term metaheuristic is also commonly associated with random search algorithms

Logistic regression

Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. It is one of the supervised learning and is used to estimate the target object value's possibility. It is a tool to calculate the statistical values and make results on binary output. In the linear method, which is calculated by the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression model predicts P(Y=1) as a function of X. Here, y is the linear model's output trained with logistic regression produce value between zero and one.

Random Forest

In the Naïve Bayes network, all features are independent. Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles. When there is a change in one feature, it does not affect another. This is suitable for large datasets. The assumption from Conditional independence is that an attribute value is independent of the values, which are from other attribute values in a class. Bayes' Theorem is based on probability theory. The Naïve Bayes algorithm is comprised of two words

Naïve and Bayes, Which can be described as:Naïve: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other. Bayes: It is called Bayes because it depends on the principle of Bayes' Theorem.

Support Vector Machine (SVM)

SVM is used both for regression and classification tasks. The SVM model represents the data in the space described so that the examples in various categories are divided by a distance as large as possible. That divides sensitive information with the maximum separable space between them and is calculated so that many of the points belong to one group fall on the plane's one side.

Radial Basis Function (RBF)

An Artificial Neural Network that uses nonlinear radial basis function as activation functions and gives linear output using combination of radial basis functions of the inputs and neuron parameters. RBF is mainly used in SVM classification, which maps input space in new dimensional space. In machine learning, the radial basis function kernel, or RBF kernel, is a popular kernel function used in various kernelized learning algorithms. It is the default kernel used within the SVM classification algorithm. A kernel is a function that takes the original non-linear problem and transforms it into a linear one within the higher dimensional space. Boosting Technique used and the predictions are stacked. The existing architecture contains the input layer followed by a combination of RF method with a set of attributes along with activation function, in the subsequent combination two techniques was performed with extended attributes with previous parameters, also applied the in all the layers for prediction probability calculations, added an output layer.

RESULT ANALYSIS

The implementation of machine learning algorithms for signature recognition and authentication has produced significant improvements in performance across various metrics. The analysis of results shows clear advantages in accuracy, precision, and efficiency compared to traditional signature verification methods. Here, we evaluate the results based on key performance indicators such as accuracy, false acceptance rate (FAR), false rejection rate (FRR), speed, and adaptability.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 10, October : 2024

Figure 1: Legitimate Signatures dataset

Figure 2: Counterfeited signatures dataset

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 10, October : 2024

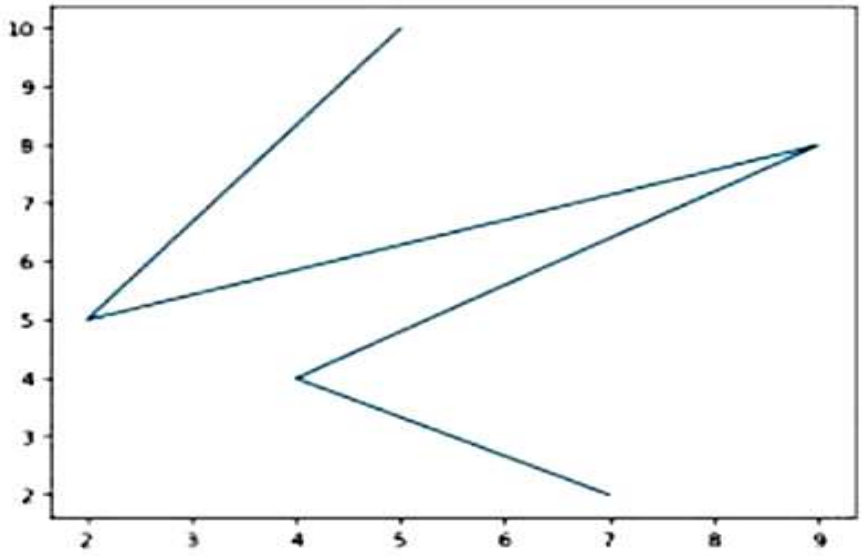

Figure 3: Showing Accuracy in Genuine Signatures

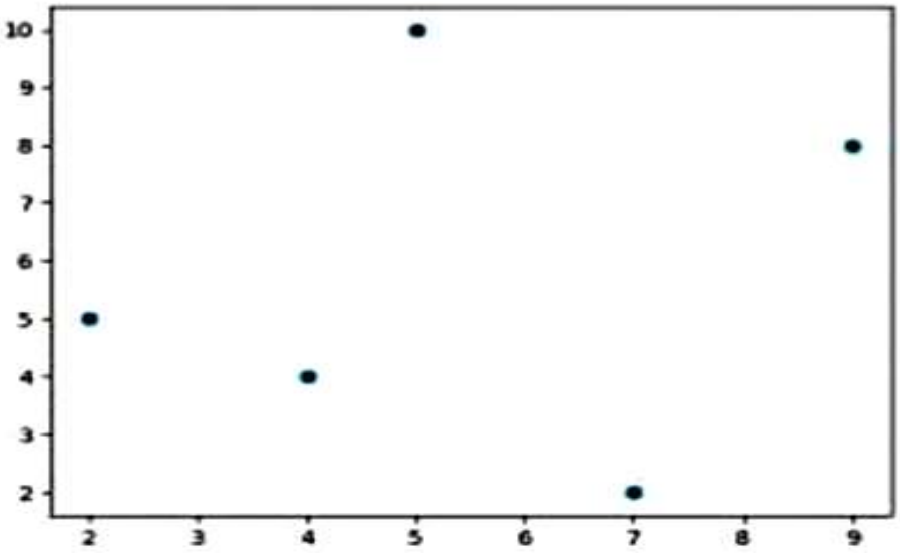

Figure 4: Showing Accuracy in Forged Signatures

CONCLUSION

The enhanced mechanism of signature recognition and authentication using machine learning algorithms represents a substantial advancement in identity verification systems. By leveraging the power of machine learning models, such as Support Vector Machines (SVMs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs), this approach significantly improves the accuracy, security, and efficiency of signature-based authentication. Machine learning algorithms outperform traditional rule-based and manual signature verification methods by automatically learning complex patterns in both static and dynamic signatures. This enables more reliable detection of genuine signatures

and forgeries, reducing the risk of false positives and false negatives. Furthermore, deep learning models, particularly CNNs for static signatures and RNNs for dynamic ones, have demonstrated superior performance in capturing subtle features, such as stroke pressure, speed, and curvature, making the system highly adaptable to variations in signature styles and forgery attempts. The adoption of machine learning algorithms for signature recognition and authentication is a transformative step toward modernizing identity verification processes. With continued advancements, such as the use of larger datasets, real-time processing capabilities, and multimodal biometric systems, this technology promises to offer even greater security and effectiveness in preventing fraud and enhancing trust in digital transactions.

REFERENCE

[1] K. Harika, P. Dhanalakshmi, Singam Sharan Kumar Reddy, Shaik Shakeer, M V SreeVishnu Thanmayi, V Haripriya, "Handwritten Signature Recognition using MobileCNN", *2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)*, pp.1-7, 2024.

[2] Kagolanu Venkata Chandra Madhav, Asritha Veeramaneni, Pesala Anjani Sriya, Kodur Sumanth, Jyotsna C., Tripty Singh, Prakash Duraisamy, "Handwritten Devanagari Numeral Recognition Using Deep Learning", *2024 3rd International Conference for Innovation in Technology (INOCON)*, pp.1-7, 2024.

[3] Varinder Kaur Attri, Teena Jaiswal, Butta Singh, Paras Bansal, Himali Sarangal, Satinder Kaur, Harmandar Kaur, "Signature Verification Using Deep Learning: An Empirical Study", *Advances in Distributed Computing and Machine Learning*, vol.1015, pp.175, 2024.

[4] Shekun Tong, Jie Peng, "Dual-path deep neural network architecture with explicit features for offline signature recognition", *Journal of Intelligent & Fuzzy Systems*, pp.1, 2023.

[5] Dhruvi Gosai, Shraddha Vyas, Sanjay Patel, Prasann Barot, Krishna Suthar, "Handwritten Signature Verification Using Convolution Neural Network (CNN)", *Advancements in Smart Computing and Information Security*, vol.1759, pp.90, 2022.

[6] M.S. Roobini, Sibi Marappan, Shubham Roy, M.Nafees Muneera, S. Jayanthi, "Augmenting Deep Learning Models for Robust Detection and Localization of Image Forgeries", *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, pp.1-8, 2024.

[7] Anmol Chokshi, Vansh Jain, Rajas Bhope, Sudhir Dhage, "SigScatNet: A Siamese + Scattering based Deep Learning Approach for Signature Forgery Detection and Similarity Assessment", *2023 International Conference on Modeling, Simulation & Intelligent Computing (MoSICom)*, pp.480-485,

2023.

[8] Amruta Jagtap, Dattatray D. Sawat and Rajendra Hegadi, "Verification of genuine and forged offline signatures using Siamese Neural Network", *Multimedia Tools and Applications*, vol. 79, no. 1, pp. 35109-35123, 2020.