CIND 860: Advanced Data Analytics Project

Section: DJH

Student Name: Muhammad Zaka Shaheryar

Student #: 500648718

Supervisor Name: Tamer Abdou

Date of Submission: Feb 25th, 2024

Project Title: Analysis of Sign language MNIST using Deep learning

Project Deliverable: Literature Review, Data Description, and Approach

Github Link: https://github.com/Zaka123456/CIND860

1. ABSTRACT

For CIND 860 Capstone project, the dataset chosen initially is changed to “Sign Language MNIST (https://www.kaggle.com/datasets/datamunge/sign-language-mnist). This is done to apply deep algorithm effectively on the dataset. Dataset is taken from kaggle and it contains two csv files. First is sign_mnist_train which consist of 27455 rows and 785 columns. Second is sign_mnist_test which has 7172 rows and 785 columns. Each column is a pixel while each row is a image. In total there are 34627 images of hand gestures.

“American Sign Language(ASL) is a complete, natural language that has the same linguistic properties as spoken languages, with grammar that differs from English. ASL is expressed by the movement of hands and faces. This dataset consists of 27,455 images of hand signs, each image is of 28 x 28 size and in grayscale format. The dataset format is patterned to match closely with the classic MNIST. Images in the dataset belong to a label from 0–25 representing letters from A-Z(but no cases of 9=J or 25=Z as they involve hand motion). The training data(27,455 cases) and the test data(7,172 cases) are approximately half the size of standard MNIST but otherwise similar to a header row of the label, pixel1, pixel….pixel784. The original hand gesture image data represented multiple users repeating gestures against different backgrounds. The Sign Language MNIST data came from greatly extending the small number (1704) of the color images included as not cropped around the hand region of interest.” (Singh, A). Table 1 below provide the research questions for the project.

Table 1: Research Questions

No.	**Research Question (RQ)**
1	Is CNN's performance practical when compared to traditional machine learning techniques such as Random Forest, SVM or KNN?
2	What models of the CNN is the most efficient to use. How the CNN's performance compared to other models like MobileNet or Deep CNN?
3	How to use use Bayesian network for causation? Can we find real cause of label?

2. INTRODUCTION

Ability to listen and hear enables effective communication. But people who suffer from hearing impairment or hearing loss face difficulty in this regard. According to WHO, there are 300 million deaf and 1 million mute [Paper 4]. Therefore, there is a strong demand for a proper way of communicating with these people. Sign language is one of the ways to solve this problem. A deaf-mute is person who is either deaf or both deaf and dumb. [Paper 5]. Sign language can help these people to communicate with each other and other people. Hearing impairment can be caused due to age, genes, noise exposure, infections, ear injury, or medicine. [Paper 5].

Sign language recognition has earned considerable attention due to the large global deaf and mute population. There are 144 different sign languages based on the regional languages [Paper 3]. This project will focus on American Sign Language (ASL).

This project's objective is to perform a comprehensive examination of Convolutional Neural Networks (CNNs) on the Sign Language MNIST dataset. In addition to that, various CNN models like MobileNet, and deep CNN will be explored and compared with traditional classifiers like SVM, KNN etc. The evaluation measures will be accuracy, precision, and recall. The dataset comprises 60,000 training and 10,000 testing grayscale images distributed over 24 ASL classes or alphabets [Paper 5].

This study offers a detailed description of the dataset characteristics and outlines the systematic evaluation methodology. Furthermore, it provides conclusive insights into the promise of CNNs over machine learning for the vitally important domain of sign language recognition. By leveraging standardized protocols and Python-based experiments with Keras and TensorFlow, this project aims to facilitate real-time sign language recognition (SLR) systems.

The tools that will be used are Python libraries used for deep learning and traditional machine learning algorithms. The systematic data analysis process approach will be used for the project. After data selection, initial analysis will be carried out followed by the exploratory analysis (EDA). Then experimental design and model building will be carried out. Finally performance evaluation will be done together with project limitations and conclusion.

3. LITERATURE REVIEW

The motivation behind this study is to advance American Sign Language (ASL) recognition technology by analyzing the Sign Language MNIST dataset through machine learning and deep learning approaches. This effort aims to improve communication tools for individuals with hearing impairments by enhancing the accuracy of ASL recognition systems. By exploring various algorithms and methodologies, the study seeks to provide valuable insights for educators, technologists, and the broader community, contributing to the development of more effective and accessible ASL educational and communication aids.

This study provides a detailed analysis of the Sign Language MNIST dataset, focusing on enhancing American Sign Language (ASL) recognition through machine learning (ML) and deep learning techniques. By examining 34,627 grayscale images representing ASL alphabets (excluding dynamic letters), our research explores the significance of individual pixels in classification, the effectiveness of various ML algorithms, and the potential of ensemble models for improved accuracy. Initial exploratory data analysis (EDA) revealed equitable label distribution and identified influential pixels through filter-based, wrapper-based, and hybrid methods, enhancing model precision. Comparative analysis between traditional ML algorithms and deep learning models, especially Convolutional Neural Networks (CNNs) will be used to validate the hypothesis that CNNs, despite being time-intensive, offer superior accuracy. Our findings contribute to ASL recognition technology, providing insights for educators, technologists, and the broader community to aid communication for individuals with hearing impairments, emphasizing the importance of feature selection, model choice, and the innovative integration of methodologies for advancing ASL recognition systems.

The questions to considers for literature review are provided in Table 2. Six papers are reviewed and key take-aways and how it can be useful to this project is provided below.

Table 2: Questions to consider for Literature Review

No.	Literature Review Question
1	What do you already know about the topic?
2	What do you have to say critically about what is already known?
3	Has anyone else ever done anything exactly the same?
4	Has anyone else done anything that is related?
5	Where does your work fit in with what has gone before?
6	Why is your research worth doing in the light of what has already been done?

The literature review on enhancing ASL recognition with machine learning and deep learning delves into the substantial groundwork laid by prior research, highlighting significant strides in algorithm application and dataset utilization. It critically acknowledges the existing gap in developing a universally accurate, real-time recognition system, despite notable efforts. This review positions the novel contributions of the current study within the context of related endeavors, underscoring its potential to address unresolved challenges in gesture variability and real-time processing. It justifies the research's relevance by emphasizing its aim to refine communication tools for the hearing impaired, thereby enriching the technological landscape and making a meaningful societal impact.

The following six papers present advancements in sign language recognition using machine learning and deep learning. They explore various methodologies, including CNNs, MobileNet, and machine learning classifiers, achieving accuracies ranging from 91.7% to 99.95% in recognizing sign languages and gestures. These studies demonstrate the potential of integrating advanced computational techniques for enhancing ASL recognition systems, contributing to improved communication tools for individuals with hearing impairments and educational applications. Their collective findings underscore the importance of optimizing recognition accuracy and real-time processing capabilities in developing accessible and effective ASL interpretation technologies.

Paper 1:

The study titled "Mediapipe and CNNs for Real-Time ASL Gesture Recognition" (Rupesh Kumar) describes a real-time system for ASL movement identification using modern computer vision and machine learning approaches, achieving a remarkable accuracy of 99.95%. This paper will serve as a main papaer for the literature review, highlighting the potential of Mediapipe and CNNs in enhancing communication for people with hearing impairments through real-time sign language recognition.

The paper presents a highly effective American Sign Language (ASL) recognition model utilizing Convolutional Neural Networks (CNN) for feature extraction and classification, achieving an impressive 99.95% accuracy. Key takeaways include the use of the Mediapipe library for real-time hand tracking and feature extraction, the development of a robust classification system capable of recognizing all ASL alphabets, and the potential extension of this technology to other sign languages. This research is particularly useful for the project as it demonstrates a scalable, highly accurate approach to ASL recognition, suggesting that integrating Mediapipe and CNNs could significantly enhance real-time ASL interpretation systems, improving communication tools for individuals with hearing impairments.

Paper 2

The paper "Sign Language Recognition Using Convolutional Neural Networks" explores the use of CNNs for recognizing Italian sign language gestures with a focus on achieving high accuracy and generalization across different users and environments. Utilizing Microsoft Kinect, the study emphasizes automated feature extraction through CNNs, managing to recognize 20 gestures with a cross-validation accuracy of 91.7%. The approach demonstrates the potential of CNNs in sign language recognition, offering valuable insights for projects aiming to enhance communication for the Deaf community. This research could guide the development of more accessible and effective ASL recognition systems, highlighting the importance of deep learning in automating sign language interpretation.

Paper 3

The paper, "Arabic Sign Language Recognition Using Convolutional Neural Network and MobileNet," focuses on developing a CNN-based model to recognize Arabic sign language with a 94.46% accuracy using the ArASL2018 dataset. This approach demonstrates an improvement over prior models in recognition accuracy. For the project, this study underscores the effectiveness of combining CNNs with MobileNet for sign language recognition, offering a potential framework for enhancing ASL recognition systems by leveraging similar methodologies to achieve high accuracy and performance in real-world applications.

Paper 4

The paper investigates Sign Language Recognition (SLR) using various machine learning algorithms and dimensionality reduction techniques to improve model training speed and classification accuracy. Key takeaways include the effectiveness of PCA and manifold learning in reducing data dimensionality, the superior performance of CNNs in image-based SLR, and the comparative analysis of machine learning models like RFC, KNN, GNB, SVM, and SGD in recognizing sign language gestures. This research could significantly benefit the project by providing insights into optimizing SLR systems for enhanced accuracy and efficiency, especially in real-time applications.

Paper 5

The paper, "Classification of Sign Language Characters by Applying a Deep Convolutional Neural Network," presents a deep CNN model to identify American Sign Language alphabets from the Sign Language MNIST dataset, achieving a 97.62% accuracy. This study outperforms previous approaches by integrating advanced CNN architectures, highlighting the potential of deep learning in sign language recognition. For the project, this research provides a robust framework for improving ASL recognition accuracy, showcasing the effectiveness of deep CNNs in handling complex image classification tasks within sign language interpretation.

Paper 6

The paper focuses on object recognition using NAO robots, emphasizing typed text, color, shape recognition, and sign language gestures with accuracies around 90%. Machine learning classifiers tested on the MNIST dataset showed 87%-98% accuracy. This research contributes to children's visual learning enhancement, suggesting future improvements for broader image applicability. The study highlights the potential of integrating advanced object recognition and machine learning techniques in educational settings, offering insights into developing more interactive and effective learning tools through robotics.

4. DATA DESCRIPTION

Dataset Description

The Pandas Profiling Report for combined dataset provides a comprehensive analysis of a dataset with 34,627 observations and 21 variables, detailing aspects like missing data, duplicates, and feature correlations. Key findings include a 0% missing data ratio, less than 0.1% duplicate rows, and high correlations among certain pixel values, indicating redundancy or strong dependencies between features. The analysis also identifies outliers and zeros in specific pixel values, crucial for preprocessing steps in machine learning projects.The report's insights on attribute types, outliers, and the dataset's memory footprint can guide data preprocessing and feature engineering, enhancing model accuracy and efficiency for projects focused on image classification or similar tasks.

5. APPROACH

Steps	Description
1. Choosing Dataset and Theme	Selection of the Sign Language MNIST dataset from Kaggle for the project theme.
2. Cleaning/Preparing Data	Converting labels to categorical for classification, performing Shapiro-Wilk test, preparing and revising EDA report, and reviewing existing literature on the dataset.
3. Initial Problem Analysis	Writing a literature review with a focus on the research questions rather than machine learning techniques.
4. Exploratory Data Analysis (EDA)	Describing data, identifying missing values and outliers, analyzing attribute types, conducting descriptive statistics, box plotting for distribution analysis, and correlating attributes.
5. Feature Selection	Using a hybrid wrapper and filter approach for selecting the most predictive pixels, with a percentile-based cutoff.
6. Experimental Design and Cross-Validation	Statistically selecting 4-5 classification algorithms, including CNN, and explaining the choice of classification algorithms.
7. Predictive Modelling	Evaluating models based on accuracy, precision, recall, and a 25x25 confusion matrix; assessing models on stability, efficiency, complexity, and reliability; and performance evaluation.
8. Conclusion and Recommendation	Concluding the analysis and making recommendations.
9. Limitations, Future Direction	Discussing the limitations of the current study and suggesting directions for future research.

6. REFERENCES

[0] Singh, A. (2020, July). Sign language MNIST problem: American Sign Language. Retrieved from https://medium.com/@abhkmr30/sign-language-mnist-problem-american-sign-language-48896ea960e0

[1] Yadav et al. (2019). Recognition of Static Hand Gestures in Indian Sign Language. *Journal of Network Communications and Emerging Technologies.

[2] Pu et al. (2019). Iterative Alignment Network for Continuous Sign Language Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[3] Leal et al. (2018). A Real-Time Deep Learning Sign Language Translation System. *2018 International Conference on Computational Science and Computational Intelligence (CSCI).

[4] Cui et al. (2019). A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training. *IEEE Transactions on Multimedia*, 21(7), 1880-1891.

[5] Bilgin & Mutludoğan (2019). American sign language character recognition with capsule networks. *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 1-6.

[6] Hasan et al. (2020). Classification of Sign Language Characters by Applying a Deep Convolutional Neural Network. *2nd ICAICT.

[7] Kayed, Anter & Mohamed (2020). Classification of garments from fashion MNIST dataset using CNN LeNet-5 architecture. *2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE).

[8] OpenAI. (2023). *ChatGPT* (GPT-4 model) [Large language model]. https://chat.openai.com/chat