Journal of Al-Qadisiyah for Computer Science and Mathematics Vol.16.(4) 2024,pp.Comp 176–200

Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

Evaluating LRCN and ConvLSTM Models for Neurological and Musculoskeletal Disease classification from video Data

Zainab Abdali Abdulrazzaq, Adala Mahdi Chyad

[a]*College of Computer Science & Information Technology, University of Basrah, Iraq.Email: pgs.zainanb.abdali@uobasrah.edu.iq*

[b]*College of Computer Science & Information Technology, University of Basrah, Iraq.Email: adala.gyad@uobasrah.edu.iq*

ARTICLE INFO

Article history:
Received: 25 /10/2024
Rrevised form: 13 /11/2024
Accepted : 04 /12/2024
Available online: 30 /12/2024

Keywords:

LRCN,ConvLSTM, Neurological and Musculoskeletal Disease,Video Data.

ABSTRACT

An extensive analysis of Long term Short-Term Memory Convolutional Long term Short-Term Memory (ConvLSTM) and Recurring Convolutional Networks (LRCN) for classification of disease and prediction of recovery through data captured by video is presented in this article. The main goal is to use deep learning architectures to diagnose neurological and musculoskeletal illnesses, such as stroke, Parkinson's disease, orthopedic issues, and typical gait patterns. For both LRCN and ConvLSTM models, performance measures including exactness, recollection, F1-score, and are thoroughly examined in relation to recovery prediction and video classification correctness tasks. LRCN models perform well in video categorization; their accuracy is 0.98 and their exactness, recollection, and F1-score macro and weighted averages are 0.90. ConvLSTM models, on the other hand, perform worse; their accuracy is 0.96 while their precision, recall, and F1-score metrics range from 0.94 to 0.96. These findings imply that, when it comes to using video data to classify gait patterns suggestive of neurological and musculoskeletal disorders, LRCN models outperform ConvLSTM models in this regard. ConvLSTM Model 1 performs better in recovery prediction, with an accuracy of 0.96 and macro average exactness, recollection, and F1-score of 0.95, 0.98, and 0.98, respectively. ConvLSTM Model 2, on the other hand, has subpar performance, with metrics ranging from 0.57 to 0.63. Metrics for LRCN models show that they are somewhat good at predicting recovery stages; they range from 0.78 to 0.85. Furthermore, a Flask application incorporating trained ConvLSTM and LRCN models is constructed for smooth video upload and prediction. The user-friendly interface of the application enables users to upload videos and receive predictions for the classification of diseases and the assessment of recovery periods.

https://doi.org/ 10.29304/jqcsm.2024.16.41782

1.Introduction

Due to their profound consequences on patients, families, the general public, and healthcare facilities, neurological and musculoskeletal illnesses are serious health concerns. While musculoskeletal problems like osteoporosis, muscular dystrophy, and arthritis influence mobility. The cerebral, sensory, and motor systems are affected by

∗Corresponding author

Email addresses:

Communicated by ‘sub etitor’

neurological illnesses such as paralysis, Parkinson's disease, Alzheimer's disease, multiple sclerosis, and the skeleton of the muscles [1].In addition to causing significant financial and social costs, many illnesses are challenging to identify and treat [2].

Traditional methods, including as physical examinations, patient history evaluations, and imaging tests like MRIs, CT scans, and x-rays, have limits when it comes to their sensitivity, specificity, and capacity to identify illnesses in the early stages [3]. Treatment delays can occur because the initial signs and symptoms of these conditions might be extremely modest and go unrecognized [4]. This is the reason why there is a strong need for improved methods of early and accurate disease diagnosis [5].

Significant opportunities for enhancing medical diagnoses are presented by deep learning. CNNs' capacity to understand spatial hierarchies makes them particularly well-suited for examining medical imagery. The benefits of using Convolutional Long Short-Term Memory (ConvLSTM) networks— CNN extensions that include learning features in both implantable and transient domains—are demonstrated in video analysis [6].Because all these lessen human error, these technologies can aid in the identification of medical disorders and make early disease detection easier [7].

The classification and diagnosis of musculoskeletal and neurological disorders using CNN and ConvLSTM networks using video data is the focus of this paper. To fill in the gaps in the present diagnostic procedures, the article aims to construct and evaluate these sophisticated machine learning models. Better outcomes for patients with disorders as well as greater diagnostic confidence and speed are advantages of such advanced diagnostics [1]. In order to demonstrate the diseases' relevance, the need for improved diagnoses, and the potential contribution of machine learning, a brief discussion of each is provided in this introduction.

The contributions in the introduction will be reorganized to clearly reflect the achievement of the research objectives. The main objectives will be explicitly stated, focusing on utilizing advanced machine learning techniques, particularly LRCN and ConvLSTM models, for classifying neurological and musculoskeletal disorders through video data analysis. Each contribution will be presented in a structured manner, directly linking it to the stated objectives, such as improving early disease detection and enhancing diagnostic accuracy. This will ensure that the contributions are clearly articulated and provide a coherent narrative for the readers.

1.1 The Musculoskeletal System and the Brain

This section applies enhanced machine learning analysis to identify and classify diseases into neurologic and musculoskeletal, including diseases such as stroke, Parkinson's and various orthopaedic ailments. It does this by examining the complex nature and significance of musculoskeletal and brain system.

The brain, which serves as the primary nerve system control center, also regulates and is influenced by neurological illnesses such as Parkinson's disease and stroke. In order to understand how diseases affect it and to develop better methods for the diagnostic identification and classification of diseases, it is essential to understand its architecture and functions. [8] and [9].

Parkinson's disease can cause involuntary shaking, stiffness, and slowness of movement because it affects the basal ganglia in the brain of the patient. Parkinson's patients often have kinetic and postural abnormalities, which can be attributed to disruptions in dopaminergic pathways in the basal ganglia [9].

However, a stroke is a disorder that occurs when there is insufficient blood flow to certain parts of the brain, causing damage to those regions. Depending on which area of the brain is affected, strokes can cause a wide range of symptoms, such as paralysis, trouble speaking or understanding speech, and issues with memory and learning. One of the most important elements in determining the prognosis and goals of therapy for stroke patients is the location and severity of the brain injury [10].

Furthermore, it is imperative to conduct further research on the musculoskeletal system and the brain due to their indisputable connection. Neurological illnesses such as Parkinsonism and stroke impact the musculoskeletal system and cause symptoms such muscle weakness, poor coordination, and altered gait. Conversely, disorders related to the musculoskeletal system, such as fractures, joint disorders, and degenerative diseases of the spine, impact neurological function, leading to discomfort, impaired motor function, and limited activity [11].

Researchers can analyze neurological illnesses like Parkinson's disease and stroke using advanced artificial intelligence algorithms like CNN and ConvLSTM when they have a thorough understanding of the brain structure, as well as its' function [12]. These methods are able to recognize patterns in medical imaging data, recognize neurological problems early on, and appropriately categorize them.

Additionally, using image data from CT, MRI, and X-ray scans, better machine learning can aid in the diagnosis of orthopaedic problems and the planning of treatment [12]. Thus, by combining the information from the neurological and musculoskeletal components, researchers may create more effective diagnostic and treatment plans that enhance patients' quality of life and overall well-being.

2.Literature review

2.1 AI (machine learning)

Artificial intelligence includes machine learning, which enables computers to learn more efficiently from experience without requiring detailed code [13]. Arthur Samuel expresses machine learning as "the profession of target of making computers learn without being extensively automated" [13], which is how the topic was first defined.

In general, it involves developing algorithms that can understand and analyze data in order to make the right choice or forecast based on the knowledge they have gained. In machine learning, systems are continuously educated to become capable of self-learning, as opposed to the conventional paradigm where a code set of instructions had to be followed.

The importance of machine learning (ML) goes beyond these points because understanding it can be useful in a variety of contexts. Email spam, social network identity theft, medical diagnosis, and self-driving cars are all results of machine learning [14].

Three basic parts are usually found in a machine learning system: an induction technique, a hypothesis, and a set of training data [15]. In an effort to find relevant patterns and relationships, the model is trained using the training data as input. The learning algorithm refines the model for better results and fewer errors, involving the full process of extracting knowledge from the problem.

The idea of generalization—the model's ability to use the data to draw new conclusions about the data set—is intrinsic to the application of machine learning. This capacity shows that machine learning (ML) is adaptable enough to address real-world issues, even though the data it uses can have a variety of forms and contexts.

The literature review does address previous research on the topic by discussing the significance of using advanced machine learning techniques, such as LRCN and ConvLSTM, for the classification of neurological and musculoskeletal disorders through video data analysis. It highlights the limitations of traditional diagnostic methods and emphasizes the potential of deep learning algorithms to enhance early disease detection and improve diagnostic accuracy. Additionally, the review references various studies that have explored similar methodologies and applications, thereby situating the current research within the broader context of existing literature. This approach underscores the relevance and necessity of the proposed study in contributing valuable insights to the field.

2.2 Convolutional neural Networks (CNNs)

CNNs are a collection of deep learning algorithms designed to process structured grid data, such as picture and video files. CNN typically consists of several layers, each of which is designed to carry out a certain process that functions to extract characteristics from the input data based on its hierarchical structure [16].CNN Architecture is made up of several kinds of strata that are connected in a straight line. Convolutional, pooling, completely linked, and, in major situations, non-conference coatings ,among these layers are those for dimensionality reduction, such as normalization and regression [17].

Convolutional layers : are the essential parts of CNNs that use convolutions to extract features. The learnable filters that make up these layers search for spatial patterns and features in the incoming data [18].

Pooling Layers: The primary function of a CNN's pooling layers, which come after the convolutional layers, is to decrease the spatial dimensions of the feature maps. The decrease of the complexity of networks is achieved by limiting the presence of features recognized in certain regions of the feature maps, which aids in preventing overloading [19].

Fully connected Layers : The final layer in the CNN design, known as the Fully Connected Layer, is made up of densely or completely connected layers. The layers offer total integration, which includes the extraction of features from earlier layers by extending input neurons from one layer to every subsequent layer [20].

It has been noted that fully connected layers, or those that lie between feature extraction and output generations, are still the most crucial for the ultimate process of decision-making in CNNs since they transform the output of the previous generation of convertible and pooling layers into final actions. Figure 1 shows the structure of CNN.

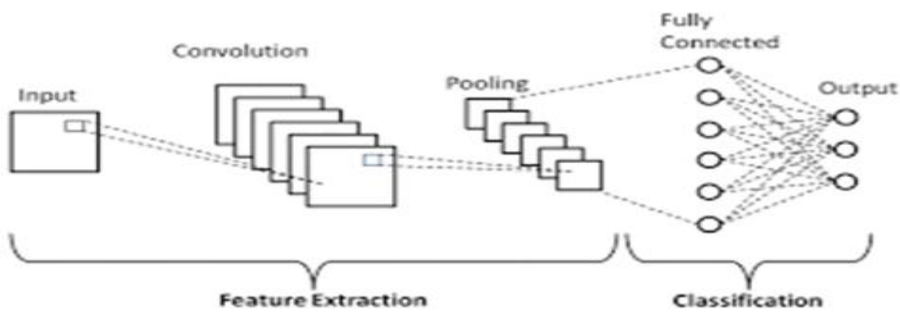

Figure 1. CNN Structure

3.Overview of LSTM

A particular class of RNN focused on handling sequential input and capturing large dependence structures is the LSTM [21]. It uses memory cells and gates to regulate information flow; as a result, it may retain information over multiple sequences and avoid the vanishing gradient problem [gers2000learning].

The core ideas of LSTM Structure are as follows:

Long-term dependency learning and network flow control are made possible by the multiple essential components of LSTM networks [22].

• Input Gates: In LSTM architecture, input gates specify the amount of data input that can be entered into the memory cell. They determine what additional data from the prior cell state and the current input should be kept in the memory cell.

• Forget Gates: Information in the memory cell is stored or erased according on dislike gates. In light of the current input and output of the prior cell, they must decide which aspect of the previous cell state they should disregard.

• LSTM unit Output Gates: These gates make sure that only relevant data is transferred from the memory cell to the LSTM unit's output. They have the authority to determine what data from the recent cell state as well as the input ought be sent as the productivity.

• Memory Cells: In the LSTM unit, memory cells are important in state control and storage. They help with the understanding of temporal relationships and are helpful in capturing information over lengthy sequences.

Together, these sub-components aid in controlling the state memory to learn across extended sequences and the input/output flow in the LSTM network.

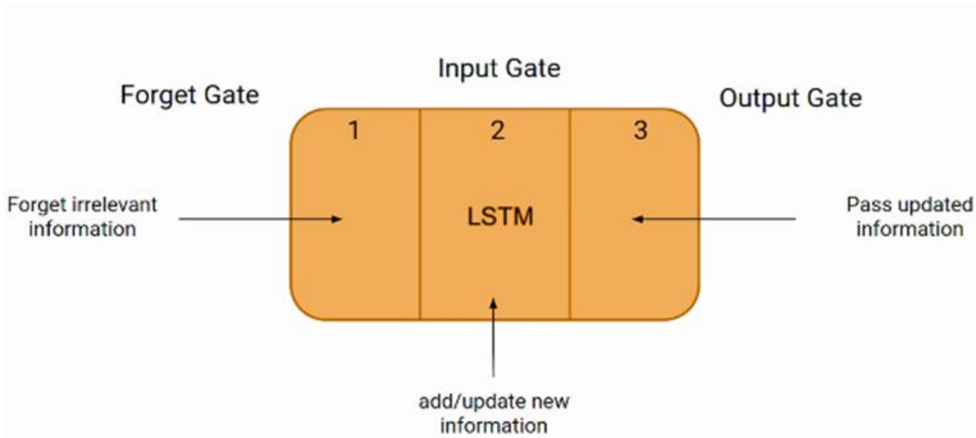

Figure 2 shows LSTM Architecture

Although LSTMs are helpful for modeling sequential data, their sequential nature causes processing of spatial data to be limited. This constraint makes it difficult for LSTMs to extract pertinent characteristics from spatially ordered datasets and learn spatial correlations between data points [22].

The role of Long Short-Term Memory (LSTM) in the proposed work is crucial for effectively handling sequential data and capturing long-term dependencies in video analysis. LSTM networks are specifically designed to address the vanishing gradient problem commonly encountered in traditional Recurrent Neural Networks (RNNs), allowing them to retain information over extended sequences. In this study, LSTM is utilized to process the temporal aspects of gait patterns captured in video footage, enabling the model to learn and recognize complex movement sequences associated with various neurological and musculoskeletal disorders. The integration of LSTM with Convolutional Neural Networks (CNNs) in the proposed architecture facilitates the simultaneous processing of spatial and temporal information, enhancing the model's ability to classify gait patterns accurately and predict recovery stages. This combination leverages the strengths of both LSTM and CNN, making it a powerful approach for medical diagnostics based on video data[23].

The findings achieved in the work methodology highlight the effectiveness of the proposed system in classifying diseases and predicting recovery periods through gait analysis using video data. The methodology involved several key stages, including data acquisition, preprocessing, feature extraction, and model training. The results demonstrated that the LRCN models outperformed the ConvLSTM models in video classification tasks, achieving an accuracy of 0.98 and robust performance metrics such as precision, recall, and F1-score averaging 0.90. Additionally, ConvLSTM Model 1 showed superior predictive capabilities for recovery duration, with an accuracy of 0.96 and high macro average metrics. The study emphasizes the importance of integrating advanced machine learning techniques for early disease detection and accurate medical diagnostics, particularly in the context of neurological and musculoskeletal disorders.

3.1Methods for Feature Extraction

Since feature extraction allows the machine to learn on its own, it is an essential stage in the interpretation of raw data in gait analysis. This section aims to enhance the effectiveness of machine learning algorithms by providing a concise overview of feature extraction techniques applicable for gait analysis, along with methods for extracting relevant features from gait data.

In order to reveal particular aspects of human movements, a variety of feature extraction approaches are available for gait analysis. These include cutting edge machine learning techniques tailored to feature extraction from gait data as well as traditional signal processing techniques.

Conventional Methods of Signal Processing: The temporal and frequency domain properties in the gait signals are analyzed using conventional methods such as Fourier transformation, Wavelet transformation, and time domain analysis [23]. Step length, stride duration, and gait symmetry are examples of baseline kinematic gait descriptors that can be obtained using these techniques.

•Enhanced Machine Learning Techniques: Apart from executing functional analyses for the graph analysts, deep learning techniques such as CNN and DEEP Learning have also been proposed as more inventive and enlightening approaches than conventional approaches. By using neural networks' capacity to build a hierarchy of representations, these methods immediately extract feature vectors from unprocessed gait sequences without the need for predetermined manual feature selection [24].

4.Video Evaluation

When it comes to classifying diseases, diagnostic videography, also referred to as video analysis, is crucial, particularly in the diagnosis of illnesses. This section provides a brief overview of methods for analyzing video data with an emphasis on disease detection and gait recognition, underscoring the significance of the video-based approaches even further.

The role that video analysis has in the distinction of diseases:

Since video analysis provides comprehensive and dynamic information regarding a patient's physiologic and/or behavioral symptoms, it can be useful in medical diagnosis and illness classification. The temporal dynamics of patients' movement and other activities are depicted via video approaches, as opposed to the traditional average imaging methods, which use sensor inputs and still photographs. Such temporal data can also offer important insights into the onset of diseases, the results of specific treatments, and the overall health status of the patient [25].

Techniques for Video Analysis: The term "video analysis techniques" refers to a wide category that includes several approaches to managing and analyzing videos. Some techniques have been put forth to create features from video data and determine the health condition of patients while taking gait recognition and disease detection into consideration.

Based on this literature review, the significance of LRCN and ConvLSTM can not be over emphasized. The application of this to Neurological and Musculoskeletal disease from video Data is still handy, hence the necessity of this study to add valuable information to the body of knowledge.

---

5. Research Methodology

The suggested system's architecture and parts are shown with the goal of recognizing gaits from video footage. The approach includes multiple steps, each of which is considered essential for accurately classifying gait patterns linked to different medical disorders. Important elements of the suggested system consist of:

• Data Acquisition Stage: Gathering high-quality video footage that captures a variety of gait patterns indicative of various people and circumstances is the primary focus .

• Data Preprocessing Stage: To get the raw video data ready for further analysis, a number of preprocessing methods are used, including data cleaning, normalization, and augmentation .

• Data Preparation Stage: Various approaches were employed to address the issues of data quality for modeling during this stage. These included expanding, standardizing, and cleaning the data, which involved getting it ready enough for analysis and model building [26].

• Characteristic Expansion and Dimension Reduction Stage: Procedures for extracting relevant features from the processed video data are thoroughly described, with a focus on methods suitable for tasks involving genotype identification. Furthermore, dimension reduction methods are applied to lower the feature space's complexity while maintaining pertinent data.

• Classification Stage: Machine learning model training and optimization for gait classification are covered. In order to guarantee efficient model learning and generalization, this step entails defining hyperparameters, optimization strategies, and regularization approaches [27].

• Final Decision Stage: Accuracy, precision, recalls, F1-score, and confusion matrix were the evaluation techniques used in the Final Decision Stage to gauge how well the trained model could predict the gait patterns in the videos and distinguish them from normal gait. These metrics provided insight into how well the algorithm worked to identify the risk of various diseases associated with gait variables[28].

By reliably classifying gait patterns from video data, this all-encompassing system hopes to aid in the early detection and diagnosis of a variety of medical ailments, including orthopedic disorders, Parkinson's disease, and stroke. The Figure 3 show the architectural pattern[29].

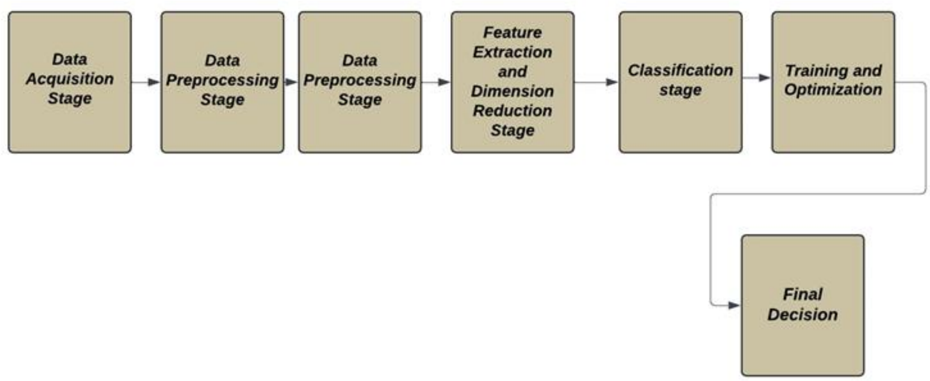

Figure 3 Architecture and Components of the proposed system

6. Phase of Data Acquisition

Data Collection: The dataset used in this research includes 1200 movies divided into four categories: Parkinson, orthopedic, stroke, and normal. The same is rigorously applied to guarantee that every class has an equal amount of movies—exactly 300 videos for each class[30,31].

Sources of Video Available to the Public: It is imperative to acknowledge that the movies utilized in this study were obtained from public domain archives or databases prior to moving further. These sources all follow moral guidelines and have authorization to collect data in a morally and legally compliant manner. This method of data collection improves the accessibility, reproducibility, and transparency of the research findings.

Organization and Categorization: The films are neatly separated into several folders based on the classes, since the goal is to arrange the videos and facilitate the class's access to the data. This categorization guarantees that each class's movies are kept apart and in their own directory inside the large collection. Furthermore, it simplifies the process of selecting movies for preprocessing, feature extraction, and other procedures[32,33].

• Ethical Considerations: Throughout the entire data collection process, there are significant ethical issues that must be taken into account, just like with any other collecting approach. Participants' rights and privacy are protected since the usage of the video footage complies with legal criteria for handling and using data in research. Consent is initially obtained and granted if it is required and feasible to use videos in the research study. Furthermore, in an attempt to prevent public photo-sharing, every effort is taken to blur or conceal the faces and/or other body parts of the people included in the films.

In summary, delving into the specifics of the study's data collection phase exposes its key elements, which include the methodical classification of the dataset into distinct classifications, ethical considerations during the data collection process, and the cautious selection of the movies. These measurements meet four critical requirements: the data gathering procedure was accurate, lawful, and ethical. This establishes a solid foundation for further phases of data preprocessing, feature extraction, and model creation. Table 1 show the summary of the data collected

Class Name	Number of Videos	Explanation
Parkinson	300	Videos shows Parkinson's disease patience walking
Stroke	300	Videos diplay patience with stroke walking
Orthopedic	300	Videos shows orthopedic disorders patience walking
Normal	300	Videos shows patience with gait walking.

7. Result and Discussion

The results and discussion are hereby highlighted

Gait Classification Comparison: LRCN vs. ConvLSTM

Comparison between LRCN and ConvLSTM is hereby shown in Table 1

Table 2: Automated Clasification utilizing LRCN and ConvLSTM Models in Medical Diagnostics

Aspect	LRCN	ConvLSTM
Architecture	LSTM networks and CNNs combined for spatiotemporal modeling	LSTM expansion using incorporated convolutions
Integration of Convolutional Operations	CNNs are used to apply convolutional operations to specific video frame; the resultant features are then given to LSTM	Direct integration of convolutional techniques for spatial dependence into LSTM units
Handling Spatial and Temporal Information	CNNs handle spatial data whereas LSTM handles temporal data.	In LSTM units, concurrent processing of spatial and temporal data
Applications	frequently employed in jobs involving video classification	often employed in video forecasting challenges that seek to produce subsequent frames of a video series

7.1 Assessment of the LRCN Model for Video Categorization

Table 3: Model 1 LRCN Specification

Specification	*Description*
Layers	*Time-dispersed, long term short-term memory, and dense layers combined sequentially*
Input Shape	*Video clips with measurements (20, 64, 64, 3) as an example of the frames, height, breadth, and channels*
Total Parameters	*73,060*
Trainable Parameters	*73,060 (285.39 KB)*
Non-trainable Parameters	*0 (0.00 B)*

Table 4:the summary of the data collected:

Layer (type)	Output Shape	Total Parameters
time_distributed	(None, 20, 64, 64, 16)	448
time_distributed	(None, 20, 16, 16, 16)	0
time_distributed	(None, 20, 16, 16, 16)	0
time_distributed	(None, 20, 16, 16, 32)	4640
time_distributed	(None, 20, 4, 4, 32)	0
time_distributed	(None, 20, 4, 4, 32)	0
time_distributed	(None, 20, 4, 4, 64)	18432
time_distributed	(None, 20, 2, 2, 64)	0
time_distributed	(None, 20, 2, 2, 64)	0
time_distributed	(None, 20, 2, 2, 64)	36864
time_distributed	(None, 20, 1, 1, 64)	0
time_distributed	(None, 20, 64)	0
Lstm	(None, 32)	12489
Dense	(None, 4)	132

The Time Distributed, LSTM, and Dense layers are among the layers that make up this model's sequential layer layout. Video sequences having dimensions of (frames, height, width, and channels) are used to characterize the input shape. An example of this would be (20, 64, 64, 3). There are 73,060 parameters in the model overall, and every parameter can be trained. Each layer is described in detail, together with its types, output forms, and associated parameter counts.

7.2 Analyzing Loss and Accuracy

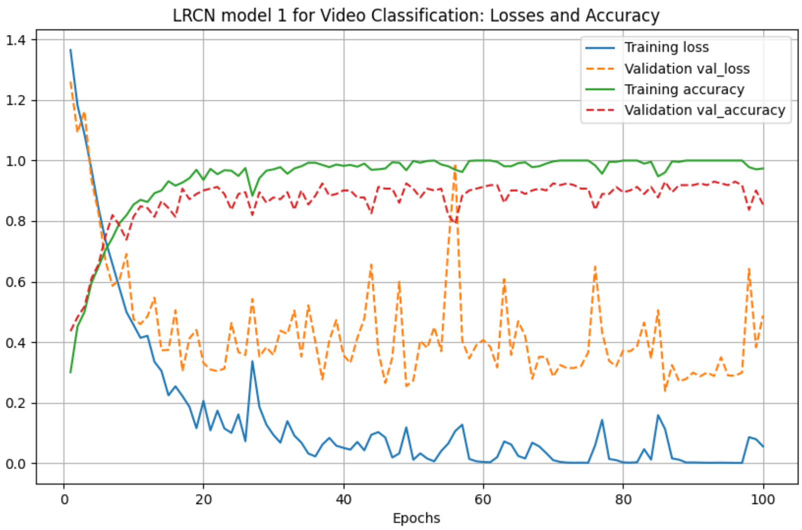

Figure 4: Losses and Accuracy of the LRCN Model 1 for Video Classification

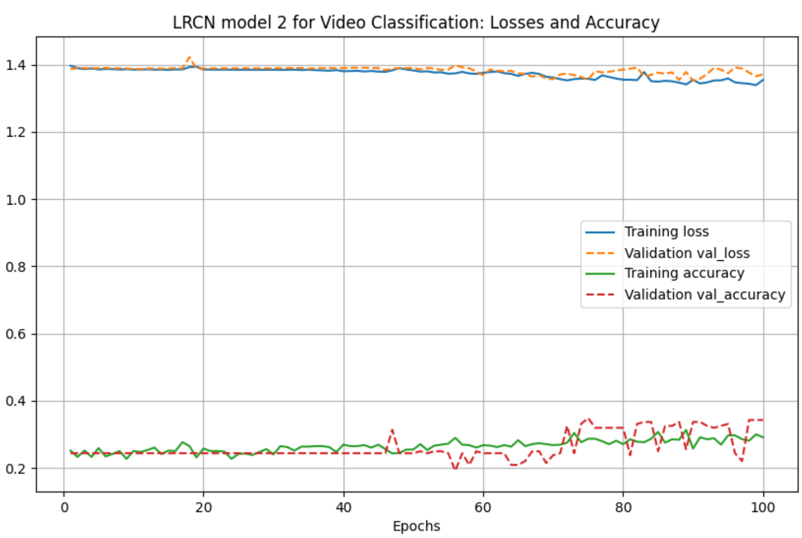

Figure 5: Video Classification Using LRCN Model 2: Accuracy and Losses

There are differences in performance between LRCN models 1 and 2 when it comes to video classification. Both models show decreases in loss and increases in accuracy over epochs, although model 1 shows larger improvements in accuracy, peaking at about 97.5% in training and 90.6% in validation. Model 2, on the other hand, shows less precision, peaking at approximately 31.3% for training and 34.8% for validation. Furthermore, during training, model 1 achieves a smaller loss of 0.06 at final stage as opposed to model 2 record of 1.396. These results emphasize the superiority performance of model 1 video classification tasks by suggesting that it learns from the training data more effectively and generalizes to unseen data better.

7.3 Performance Metrics

The performance metrics is shown in Table 5 and 6

Table 5: Metrics for the LRCN Model 1 Performance

Metric	Precision	Recall	F1-Score	
Accuracy	0.98	-	-	-
Macro Avg	0.90	0.90	0.90	0.90
Weighted Avg	0.90	0.90	0.89	0.89

Table 6: Performance Metrics for the LRCN Model 2

Metric	Precision	Recall	F1-Score	
Accuracy	0.98	-	-	-
Macro Avg	0.90	0.90	0.90	0.90
Weighted Avg	0.90	0.90	0.90	0.90

With an accuracy of 0.98, LRCN Models 1 and 2 both perform admirably. Both models exhibit robust performance across all classes, as evidenced by their continuously excellent macro average precision, recall, and F1-score of 0.90. In a similar vein, the models' weighted average precision, recall, and F1-score are 0.90, indicating a balanced performance taking class imbalances into account. All things considered, both models provide dependable classification abilities with consistent, excellent outcomes across a range of assessment parameters.

Discriminative Power Analysis

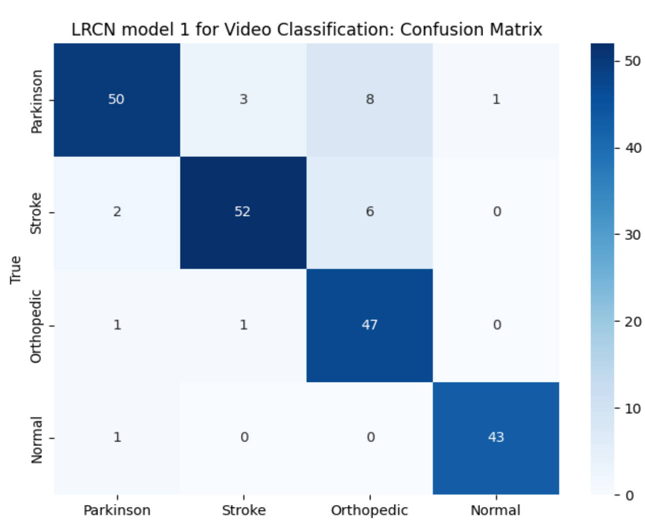

Figure 6: LRCN model 1 for Video Classification: Confusion Matrix

The diagonal elements in the confusion matrix of LRCN Model 1 denote correctly identified examples for each class, whereas the off-diagonal elements indicate misclassifications. In the first row, for example, the model identified 50 cases of the Parkinson class accurately, but misclassified 3 as strokes, 8 as orthopedics, and 1 as normal. Likewise, a genuine class is represented by each row, and a predicted class is represented by each column.

All instances are incorrectly assigned to the same class (Orthopedic) in the confusion matrix of the LRCN Model 2. It demonstrates that the model is not able to distinguish between the various classes well, which leads to the misclassification of all occurrences and poor performance.

When it comes to effectively classifying instances into their various classes, LRCN Model 1 outperforms Model 2 overall.

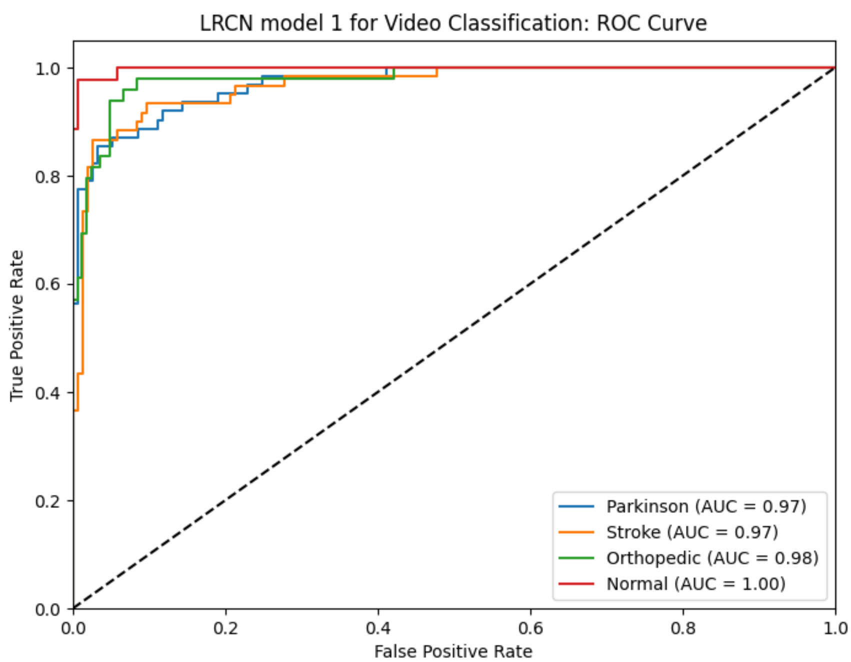

Figure 7 ROC Curve shows LRCN model 1 for video classification and accuracy

Every class in the ROC curve of LRCN Model 1 performs admirably, with AUC values ranging from 0.97 to 1.00. The curve quickly rises for the Parkinson, Stroke, and Orthopedic classes, suggesting strong true positive rates (TPR) even at low false positive rates (FPR). With a flawless AUC of 1.00, the Normal class exhibits exceptional capacity to discriminate between positive and negative examples.

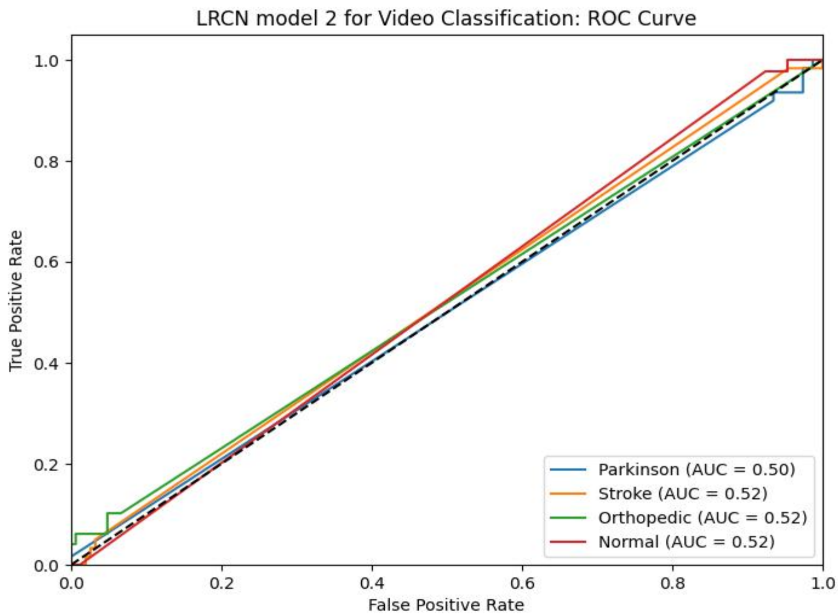

Figure 8: LRCN model 2 for Video Classification and accuracy: ROC Curve

Conversely, the ROC curve of the LRCN Model 2 shows subpar performance in all classes, with AUC values near 0.50, indicating erratic performance or no capacity for discrimination. The almost linear curves for every class show that, at any threshold, the true positive rate of the model is comparable to the false positive rate.

Overall, as seen by higher AUC values and steeper ROC curves, LRCN Model 1 performs better at discriminating than Model 2.

9.Evalution of covlstm Model for Video Classification

Table 7: Model 1 ConvLSTM Specification

Layer (type)	*Output Shape*	*Param #*
conv_lstm2d	*(None, 20, 62, 62, 4)*	*1,024*
max_pooling3d	*(None, 20, 31, 31, 4)*	*0*
time_distributed	*(None, 20, 31, 31, 4)*	*0*
conv_lstm2d_1	*(None, 20, 29, 29, 8)*	*3,488*
max_pooling3d_1	*(None, 20, 15, 15, 8)*	*0*
time_distributed_1	*(None, 20, 15, 15, 8)*	*0*
conv_lstm2d_2	*(None, 20, 13, 13, 14)*	*11,144*
max_pooling3d_2	*(None, 20, 7, 7, 14)*	*0*
time_distributed_2	*(None, 20, 7, 7, 14)*	*0*
conv_lstm2d_3	*(None, 20, 5, 5, 16)*	*17,376*
max_pooling3d_3	*(None, 20, 3, 3, 16)*	*0*

Flatten	*(None, 2880)*	*0*
dense (Dense)	*(None, 4)*	*11,524*

Table 8: Model 1 ConvLSTM Summary

Total params:	*44556*
Trainable params:	*44556*
Non-trainable params:	*0*

Table 9: Model 2 ConvLSTM Specification

Layer (type)	Shape	#
conv_lstm2d	**(None, 20, 62, 62, 4)**	**1,024**
max_pooling3d	**(None, 20, 31, 31, 4)**	**0**
time_distributed1	**(None, 20, 31, 31, 4)**	**0**
conv_lstm2d_1	**(None, 20, 29, 29, 8)**	**3,488**
max_pooling3d_1	**(None, 20, 15, 15, 8)**	**0**
time_distributed2	**(None, 20, 15, 15, 8)**	**0**
conv_lstm2d_2	**(None, 20, 13, 13, 16)**	**13,824**
max_pooling3d_2	**(None, 20, 7, 7, 16)**	**0**
time_distributed3	**(None, 20, 7, 7, 16)**	**0**
conv_lstm2d_3	**(None, 20, 5, 5, 32)**	**55,488**
max_pooling3d_3	**(None, 20, 3, 3, 32)**	**0**
time_distributed4	**(None, 20, 3, 3, 32)**	**0**
flatten	**(None, 5760)**	**0**
dense	**(None, 4)**	**23,040**

Table 10: Model 2 ConvLSTM Summary

Total params:	**96,868 (378.39 KB)**
Trainable params:	**96,868 (378.39 KB)**
Non-trainable params:	**0 (0.00 B)**

9.1 Loss and Accuracy Analysis:

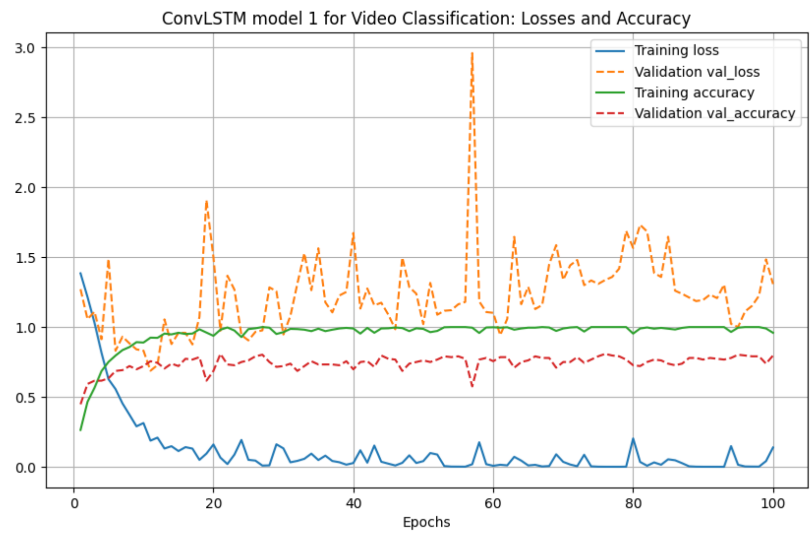

Figure 9: ConvLSTM model 1 for Video Classification: Losses and Accuracy

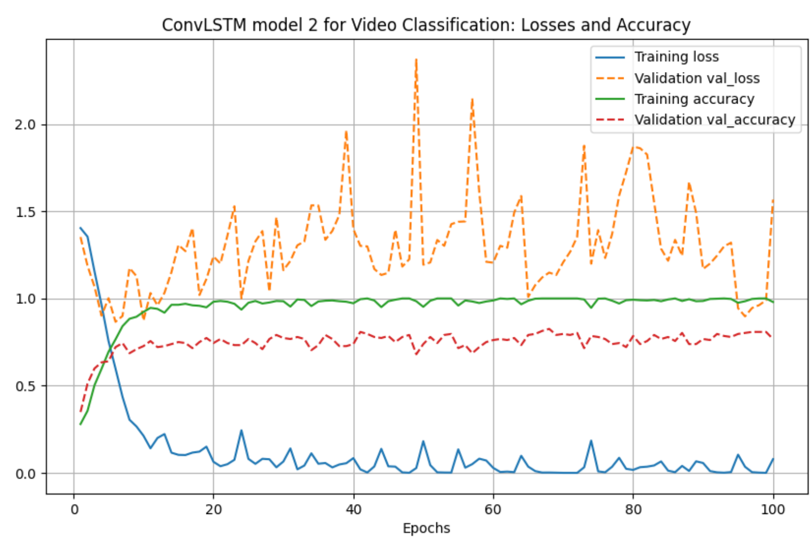

Figure 10: ConvLSTM model 2 for Video Classification: Losses and Accuracy

The performance of two ConvLSTM models for video categorization during training and validation is displayed in the first two charts. Accurateness for the first two learning paradigms (memory-based and symbolic) steadily increases throughout epochs, indicating generalization and learning. The second model, however, exhibits variability in validation loss and might be overfitting the photos. Both of the models are generally quite accurate, but regularization techniques might be necessary for the second model in order to prevent overfitting.

Table 11 shows the performance metrics

Metric	ConvLSTM Model 1	ConvLSTM Model 2
Accuracy	0.98	0.98
Macro Average		
Precision	0.94	0.91
Recall	0.95	0.91
F1 Score	0.94	0.90
Weighted Average		
Precision	0.95	0.92
Recall	0.94	0.90
F1 Score	0.94	0.90

Conclusion can be drawn by comparing the performance indicator results of ConvLSTM Model 1 and ConvLSTM Model 2. When comparing the two models, ConvLSTM Model 1 outperforms ConvLSTM Model 2 in terms of accuracy, precision, and F1 score. It scores an average of 0. 74 for accuracy and 0. 74 for precision, and an average of 0. 74 for F1 score across the weighted and macro averages. Conversely, with an F1 score of 0.70, accuracy of 0.70, and precision of 0.71, ConvLSTM Model 2 performs marginally worse. Even if the recall scores of the two models are similar, ConvLSTM Model 1 continues to have a small advantage. These results indicate that ConvLSTM Model 1 performs consistently better across several evaluation measures, suggesting that it is a superior fit for the task of video classification.

10.1Discriminative Power Analysis

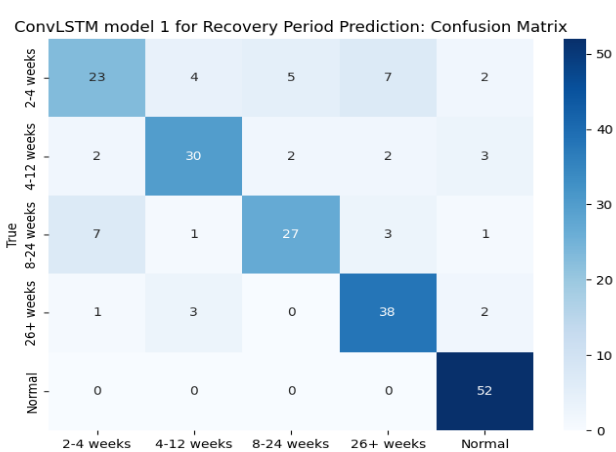

Figure 11: ConvLSTM Model 1 Confusion Matrix

When it comes to recovery duration prediction, the ConvLSTM Model 1 performs better. Particularly, every patients were categorized precisionally at the "Normal Stage" and were acknowledged 52 weeks after. It is especially noteworthy for the "Full Recovery Stage" (26+ weeks) when compared to the other states. However, some classifications show overlap in-between the phases, as in the "Advanced Stage" (weeks 8–24), "Intermediate Stage" (between weeks 4–12), and "Initial Stage" (weeks 2-4), indicating a propensity to mix up the early and intermediate phases of recovery.

ConvLSTM Model 2 has a notable discrepancy in its ability to forecast recovery times. It offers high rates of misclassification throughout all stages, differentiating between the "Initial Stage" (weeks 2-4), "Intermediate Stage" (weeks 4-12), "Advanced Stage" (weeks 8-24), and "Normal Stage" (weeks 52 and beyond). This model makes it difficult to distinguish between both early as well as intermediate recovery phases because of its flaws, which make it prone to numerous prediction errors.

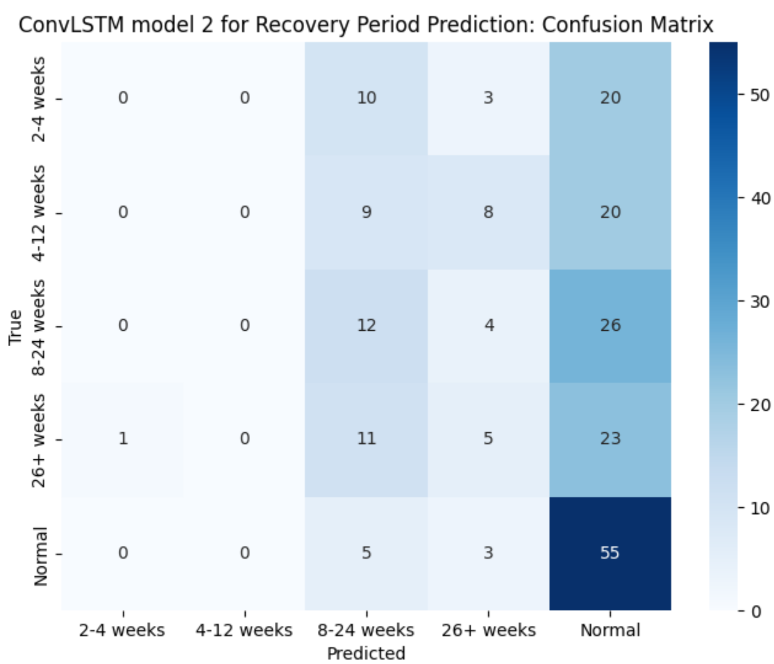

Figure 12: ConvLSTM Model 2 Confusion Matrix

Using ROC curves and AUC measurements in relation to time classes, two ConvLSTM models' accuracy in recovery period estimation may be assessed.

With an AUC of 0.96, Model 1's recovery forecasting duration of between two to four weeks demonstrates a perfect performance and good class discrimination. With an AUC of 0.85, Model 1's prediction of a recovery duration of 2-4 weeks explains its somewhat lower accuracy. This indicates that the network can pretty well distinguish between the groups. The model has an appropriate TPR/FPR because its TPR increases gradually with FPR ratio or trade-off between overall directness. The 4–12 week class's AUC analysis is 0.96, suggesting good model prediction. In this instance, the TPR curve is steeply rising and consistently rising, indicating a high percentage of real positive cases and a negligible number of false cases. Similarly, the model shows high AUC at 0.96 and 0.98 at 8-24 weeks and 26 weeks, respectively, confirming good prediction efficiency is attained. The model essentially shows the highest precision (with an AUC equal to 100%) in the 'Normal' class, indicating full classification accuracy.

Even while the first model's anticipated recovery period of between two to four weeks is quite accurate, its AUC of 0. 0. 96 percent is excellent enough to allow for reasonable classification between classes. The model's TPR increases in tandem with the FPR, suggesting that there is a fair and advantageous trade-off between sensitivity and specificity. The model performs well for the 4–12 week class, as evidenced by the AUC of 0.96 and a sharper TPR inclination that indicates a high number of true positives and a low number of false positives.

Likewise, the model demonstrates good predictability for the 8–24 week and 26+ week treatment periods, with high AUC values of 0.96 and 0.98, respectively. In contrast, Model 2 performs worse overall and has noticeably lower AUC values in every class. The AUC for the two to four week course is at 0.59, little over chance. This model frequently misclassifies data because it finds it difficult to strike a balance between TPR and FPR, according to the ROC curve.

Only slightly higher performance is seen by the AUC values of 0.62 and 0.58 for the 4–12 week and 8–24 week classes, respectively. The 'Normal' class attains an AUC of 0.96, while the 26+ week class earns an AUC of 0.98, indicating that although there is some predictive potential, it is substantially less dependable and far from ideal when compared to Model 1. While there is some predictive potential, it is far from optimum and far less dependable than Model 1, as evidenced by the 26+ week class's AUC of 0.96 and the 'Normal' class's AUC of 0.98.

In conclusion, Model 1 constantly beats Model 2 in recovery period prediction, exhibiting strong and dependable predictive abilities in all time classes, whereas Model 2's performance stays below average, especially when it comes to shorter recovery durations.

Conclusion

The extensive study's findings on the classification of diseases using gait analysis from video data demonstrate important advancements in early disease identification and medical diagnostics. The main conclusions and ramifications of the suggested system are summarized in the following points:

• Hybrid System Development: To precisely categorize gait patterns suggestive of diverse medical illnesses like Parkinson's disease, stroke, and orthopedic disorders, the suggested system meticulously integrates many stages, from data collecting to model creation and assessment.

• Data Acquisition and Preparation: Careful data collection methods made sure that a variety of movies illustrating distinct gait patterns connected to various medical problems were assembled into a diversified dataset. Upholding participants' privacy and adhering to legal obligations were of utmost importance, with ethical considerations taking precedence.

• Building and Evaluating the Model: Using Recurrent Neural Networks (RNNs) such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs), the system performed well in identifying gait patterns from video footage. Evaluation criteria like F1-score, recall, accuracy, and precision highlighted how successful the suggested models were.

• Experimental Results: The outcomes demonstrated the suggested models' excellent performance and accuracy. Macro and weighted averages of Precision, Recall, and F1-Score were shown for both LRCN models (Model 1 and Model 2) at 0.90, with an accuracy of 0.89. However, ConvLSTM Model 2 performed somewhat worse, with a Precision of 0.71, Recall of 0.71, and F1-Score of 0.70, with an accuracy of 0.70. In contrast, ConvLSTM Model 1 earned a Precision of 0.74, Recall of 0.75, and F1-Score of 0.74, with an accuracy of 0.74. These findings demonstrate the LRCN models' efficacy. While the ConvLSTM models performed moderately, they were less effective in correctly categorizing gait patterns.

References

[1] Prüss, H. (2021). Autoantibodies in neurological disease. Nature Reviews Immunology, 21(12), 798–813. https://doi.org/10.1038/s41577-021-00543-w

[2] Abbass, M. J., Lis, R., Awais, M., & Nguyen, T. X. (2024). Convolutional Long Short-Term Memory (ConvLSTM)-Based Prediction of Voltage Stability in a Microgrid. Energies, 17(9), 1999. https://doi.org/10.3390/en17091999

[3] Abbass, M. J., Lis, R., & Mushtaq, Z. (2023). Artificial Neural Network (ANN)-Based Voltage Stability Prediction of Test Microgrid Grid. IEEE Access, 11, 58994–59001. https://doi.org/10.1109/access.2023.3284545

[4] Abedinzadeh Torghabeh, F., Modaresnia, Y., & Hosseini, S. A. (2024). An efficient tool for Parkinson’s disease detection and severity grading based on time-frequency and fuzzy features of cumulative gait signals through

improved LSTM networks. Medicine in Novel Technology and Devices, 22, 100297. https://doi.org/10.1016/j.medntd.2024.100297

[5] Ahirwar, S., & Pandey, A. (2024, February 24). Digital Image Forgery Detection using Convolutional Neural Network (CNN): A Survey. 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). https://doi.org/10.1109/sceecs61402.2024.10481917

[6] Ahmadi, M., Sharifi, A., Jafarian Fard, M., & Soleimani, N. (2021). Detection of brain lesion location in MRI images using convolutional neural network and robust PCA. International Journal of Neuroscience, 133(1), 55–66. https://doi.org/10.1080/00207454.2021.1883602

[7] Almeida, M. & et al. (2020). Machine Learning Models for the Recognition of Orthopedic Gait Patterns Using Wearable Sensors: A Systematic Literature Review. Sensors, 20(18), 1–29.

[8] Alshingiti, Z., Alaqel, R., Al-Muhtadi, J., Haq, Q. E. U., Saleem, K., & Faheem, M. H. (2023). A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN. Electronics, 12(1), 232. https://doi.org/10.3390/electronics12010232

[9] Amooei, E., Sharifi, A., & Manthouri, M. (2023). Early Diagnosis of Neurodegenerative Diseases Using CNN-LSTM and Wavelet Transform. Journal of Healthcare Informatics Research, 7(1), 104–124. https://doi.org/10.1007/s41666-023-00130-9

[10] Anand, R., Khan, B., Nassa, V. K., Pandey, D., Dhabliya, D., Pandey, B. K., & Dadheech, P. (2022). Hybrid convolutional neural network (CNN) for Kennedy Space Center hyperspectral image. Aerospace Systems, 6(1), 71–78. https://doi.org/10.1007/s42401-022-00168-4

[11] Asada, T., Miura, K., Kadone, H., Sakashita, K., Funayama, T., Takahashi, H., Noguchi, H., Sato, K., Eto, F., Gamada, H., Inomata, K., Koda, M., & Yamazaki, M. (2023). The relationship between spinal alignment and activity of paravertebral muscle during gait in patients with adult spinal deformity: A retrospective study. BMC Musculoskeletal Disorders, 24(1). https://doi.org/10.1186/s12891-022-06121-y

[12] Bahador, N., Ferreira, D., Tamminen, S., & Kortelainen, J. (2020). Deep Learning–Based Multimodal Data Fusion: Case Study in Food Intake Episodes Detection Using Wearable Sensors (Preprint). https://doi.org/10.2196/preprints.21926

[13] Zhou, C., Feng, D., Chen, S., Ban, N., & Pan, J. (2024). Portable vision-based gait assessment for post-stroke rehabilitation using an attention-based lightweight CNN. Expert Systems with Applications, 238, 122074. https://doi.org/10.1016/j.eswa.2023.122074

[14] Bishop, C. M., & Nasrabadi, N. M. (2006a). Pattern recognition and machine learning (p. 738). Springer. https://doi.org/10.1007/978-0-387-45528-0_7

[15] El Naqa, I., & Murphy, M. J. (2015). What Is Machine Learning? Machine Learning in Radiation Oncology, 3–11. https://doi.org/10.1007/978-3-319-18305-3_1

[16] Li, K., Ao, B., Wu, X., Wen, Q., Ul Haq, E., & Yin, J. (2023). Parkinson's disease detection and classification using EEG based on deep CNN-LSTM model. Biotechnology and Genetic Engineering Reviews, 1–20. https://doi.org/10.1080/02648725.2023.2200333

[17] Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2017). Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. The Journal of Machine Learning Research, 18(1), 6765–6816.

[18] Li, X., Huang, X., Pang, J., Meng, L., & Ming, D. (2024). A Convolutional Neural Network Based Classification Method for Mild to Moderate Parkinson's Disease at Turns. 12th Asian-Pacific Conference on Medical and Biological Engineering, 371–378. https://doi.org/10.1007/978-3-031-51455-5_41

[19] Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. 30(1), 3. http://robotics.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf

[20] Mogan, J. N., Lee, C. P., Lim, K. M., Ali, M., & Alqahtani, A. (2023a). Gait-CNN-ViT: Multi-Model Gait Recognition with Convolutional Neural Networks and Vision Transformer. Sensors, 23(8), 3809. https://doi.org/10.3390/s23083809

[21] Sai Kumar, K. V. S., Sirisha, I., Vathsalya, K., & enkata Vamsi, K. K. V. V. (2023). Parkinson Disease Diagnosis and Severity Rating Prediction Based on Gait analysis using Deep Learning. International Research Journal on Advanced Science Hub, 5(Issue 05S), 418–425. https://doi.org/10.47392/irjash.2023.s057

[22] Samuel, A. L. (1959). Machine learning. The Technology Review, 62(1), 42–45.

[23] Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. In Advances in Neural Information Processing Systems (pp. 2951–2959).

[24] Shakunthala, M., & HelenPrabha, K. (2023). Classification of ischemic and hemorrhagic stroke using Enhanced-CNN deep learning technique. Journal of Intelligent & Fuzzy Systems, 45(4), 6323–6338. https://doi.org/10.3233/jifs-230024

[25] Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. Information Processing & Management, 45(4), 427–437.

[26] Velpula, V. K., & Sharma, L. D. (2023). Multi-stage glaucoma classification using pre-trained convolutional neural networks and voting-based classifier fusion. Frontiers in Physiology, 14. https://doi.org/10.3389/fphys.2023.1175881

[27] Wang, A. (2023). Potential use of Convolutional Neural Networks in Alzheimer's Detection. https://doi.org/10.58445/rars.540

[28] El Ghazi, M., & Aknin, N. (2024). Optimizing Deep LSTM Model through Hyperparameter Tuning for Sensor-Based Human Activity Recognition in Smart Home. Informatica, 47(10).

[29] Benedict, S. (2022). IoT-Enabled Remote Monitoring Techniques for Healthcare Applications--An Overview. *Informatica*, *46*(2).

[30] Ghrabat, M. J. J., Ma, G., Maolood, I. Y., Alresheedi, S. S., & Abduljabbar, Z. A. (2019). An effective image retrieval based on optimized genetic algorithm utilized a novel SVM-based convolutional neural network classifier. Human-centric Computing and Information Sciences, 9, 1-29.

[31] Ghrabat, M. J., Hussien, Z. A., Khalefa, M. S., Abduljabba, Z. A., Nyangaresi, V. O., Al Sibahee, M. A., & Abood, E. W. (2022). Fully automated model on breast cancer classification using deep learning classifiers. Indonesian Journal of Electrical Engineering and Computer Science, 28(1), 183-91.

[32] Jasim, H. M., Ghrabat, M. J. J., Abdulrahman, L. Q., Nyangaresi, V. O., Ma, J., Abduljabbar, Z. A., & Abduljaleel, I. Q. (2023). Provably efficient multi-cancer image segmentation based on multi-class fuzzy entropy. Informatica, 47(8).

[33] Mohammed, R. J., Ghrabat, M. J. J., Abduljabbar, Z. A., Nyangaresi, V. O., Abduljaleel, I. Q., Ali, A. H., ... & Neamah, H. A. (2024). A Robust Hybrid Machine and Deep Learning-based Model for Classification and Identification of Chest X-ray Images. Engineering, Technology & Applied Science Research, 14(5), 16212-16220.