

Salat Postures Detection Using a Hybrid Deep Learning Architecture

Khalil Ur Rehman, Sameena Javaid, Mudasser Ahmed and Shahzad Khan

Department of Computer Sciences, School of Engineering and Applied Sciences, Bahria University Karachi Campus, Sindh, Pakistan.

***Corresponding Author:** Sameena Javaid. Email: sameenajaved.bukc@bahria.edu.pk

Citation | Rehman. K, Javaid. S, Ahmed. M and Khan. S, “Salat Postures Detection Using a Hybrid Deep Learning Architecture”, IJIST, Vol. 5 Issue. 4 pp 609-625, Nov 2023

DOI: <https://doi.org/10.33411/IJIST/581>

Received | Nov 03, 2023 **Revised |** Nov 15, 2023 **Accepted |** Nov 18, 2023 **Published |** Nov 21, 2023.

Salat, a fundamental act of worship in Islam, is performed five times daily. It entails a specific set of postures and has both spiritual and bodily advantages. Many people, notably novices and the elderly, may trouble with maintaining proper posture and remembering the sequence. Resources, instruction, and practice assist in addressing these issues, emphasizing the need of prayer sincerity. Our contribution in the research is two-fold as we have developed a new dataset for Salat posture detection and further a hybrid model Media Pipe+3DCNN. Dataset is developed of 46 individuals performing each of the three compulsory Salat postures of Qayyam, Rukku and Sajdah and model was trained and tested with 14019 images. Our current research is a solution for correct posture detection which can be used for all ages. We examined the Media Pipe library design as a methodology, which leverages a multistep detector machine learning pipeline that has been proven to work in our research. Using a detector, the pipeline first locates the person's region-of-interest (ROI) within the frame. The tracker then forecasts the pose landmarks and division mask in between the ROIs using the ROI cropped frame as input. A 3D convolutional neural network (3DCNN) was also utilized to extract features and classification from key-points retrieved from the Media Pipe architecture. With real-time evaluation, the newly built model provided 100% accuracy and a promising result. We analyzed different evaluation matrices such as Loss, Precision, Recall, F1-Score, and area under the curve (AUC) to give validation process authenticity; the results are 0.03, 1.00, 0.01, 0.99, 1.00 and 0.95. accordingly.

Keywords: 3DCNN, HCI, Namaz posture, Salat gesture recognition



Introduction:

Human activities have been widely investigated using many different skills such as sensing Technologies [1][2][3], Machine Learning (ML), and recently using Vision based Deep Learning (DL). The concept of Deep Learning is the classification of a person's activity from the collected data through sensors (For example, Laser Scanner, Accelerometer, and Camera). Advances in human activity recognition have enabled a variety of applications. at a very high scope in different work areas, For Example, Medical Care, Sports Activities, Violence detection, Posture Recognition and many other fields [4].

Human Computer Interaction (HCI) is now the most powerful technology ever, influencing society and services given in a variety of applications. One of the key driving components of HCI in the recent decade has been the advent of deep learning in computer vision applications, particularly with Convolutional Neural Networks (CNNs) [5]. Deep Learning methods were used in a variety of applications, including people behavior monitoring, vehicle detection, semantic segmentation of urban environments, self-driving vehicle object detection, traffic sign gesture recognition, yoga or Salat posture detections, and semantic segmentation classification [6].

Salat is Islam's second pillar and the most important act of devotion that every Muslim does five times a day. It includes a set of postures as well as spiritual meaning. And its spiritual worth, which should be executed in a specific order. Salat with certain movements can assist develop mental well-being, increase positive personality, and improve physical health, in addition to improving the bond with Allah. Salat involves both physical effort and mental concentration [7]. Each Muslim must comply with the behaviors and set up structured movements based on the laws of Salat in performing it. With full conformity and order, Salah involves continuous gentle muscle contraction and relaxation. It contains various types of exercises for stretching and isometric contraction. For all ages & conditions, these gentle & easy exercises are appropriate.

According to the Senior Physiotherapist, Rehabilitation and Physical Medicine Institute, Al Ain Hospital, UAE, "12 'Rakah' (unit for a set of actions in a prayer) equaled 30 minutes of light exercises daily, as recommended by health experts" [8]. Other helpful activities, such as strolling to the local mosque, can be incorporated with prayer. This can assist patients in meeting the required weekly exercise levels of 30 minutes per day, five times per week. Muslim communities require a larger range of motion (ROM) because their religious activities necessitate greater flexion of the joints in the lower limbs. To fulfil this daily duty, Muslims must adopt many postures that necessitate deep knee and hip flexion [9]. Salah's varied postures were studied.

A Goniometer (an instrument that evaluates the allowable range of motion at a joint) was used to measure ROM in various investigations. Several existing strategies for recognizing human movement in daily activities have been proposed [10]. In this context, we are concerned with a specific application of human activity that is particularly important to the Muslim people around the world, namely the acknowledgment of Islamic prayer postures, commonly referred to as Salat. Another key part in Salat is calmness, as all postures should be performed carefully and slowly while reciting the Quran. There have been few attempts in the literature to explore the process of identifying Salat. Due to a lack of publicly available datasets, this topic is difficult and open to further development and usability [11].

The central motivation behind this research project is to address the reluctance of certain individuals, to seek guidance on the correct postures of Salat due to limited access to experts or resources. The project's objectives include dataset collection, leveraging this dataset for Salat posture recognition, and implementing a Salat Monitoring System to alert users about any posture errors during prayer. The system primarily focuses on three fundamental Salat postures: Qayyam, Rukku, and Sajdah. The dataset is derived from observations of

Muslims performing their daily Salat, serving as a foundational resource for posture recognition and correction. This research article aims to contribute to the improvement of Salat performance and adherence among individuals facing barriers to learning the correct postures of this essential religious practice.

Literature Review:

The study of human behaviors and postures has made tremendous advances in recent years, with numerous technologies and approaches playing critical roles. This includes the use of sensor technologies, computer vision, and, more recently, deep learning algorithms [12]. Deep learning, in particular, has transformed the classification of human actions using data from a variety of sensors such as laser scanners, accelerometers, and webcams. These advances in human activity recognition have paved the way for a wide range of applications in disciplines as diverse as medical care, sports, violence detection, and posture recognition.

Human Computer Interaction (HCI) has evolved as a significant and transformational technology with far-reaching effects on society and numerous practical uses. Deep learning has played a crucial role in HCI during the last decade, particularly in the field of computer vision, primarily through the use of convolutional neural networks (CNNs). Deep learning methods have been used in a wide range of scenarios, including monitoring human behavior, vehicle detection, semantic segmentation of urban environments, object detection in self-driving vehicles, recognition of traffic signs and gestures, and even identifying specific postures in activities such as yoga and Salat. This multidisciplinary approach to deep learning and computer vision is improving our ability to comprehend, interact with, and benefit from vast amounts of data. Despite the fact that numerous researchers have made major contributions to the subject of human activity recognition, gaps in the literature and a lack of comprehensive studies have provided momentum for additional work and inquiry in this topic. These gaps provide an opportunity for academics to go deeper, address specific issues, and make new breakthroughs, thereby improving our understanding and capabilities in this critical area of study.

In the field of Salat (prayer) posture detection hardware components like accelerometer devices and mobile devices can be of great assistance. These technologies, which are frequently integrated into mobile phones and wearable devices, aid in the collection of critical data. During Salat, worshippers take various postures such as standing (Qayyam), bowing (Ruku), and prostrating (Sujood), each with its own pattern of acceleration and orientation. The basis for posture detection is accelerometers, which measure these changes. Rahman and his fellows [13] assistive intelligence framework is intended to assist worshippers in determining the accuracy of their Salat postures by utilizing image comparison and pattern matching algorithms such as Euclidean distance, template matching, and grey-level correlation. The system provides real-time feedback by capturing images of users during Salat and comparing them to a database of correct postures, indicating whether postures are accurate and offering corrective guidance. This framework had the potential to improve Salat posture accuracy and comprehension. Similarly, several researchers [14][15] and [3] combined these devices with machine learning algorithms, they can interpret accelerometer data. Mobile devices can recognize and classify Salat postures as they are performed during prayer by training a model with known Salat postures and their corresponding accelerometer data. This real-time posture recognition can provide instant feedback, assisting individuals in ensuring that their Salat postures are correct [16]. Over several other benefits one of the benefits of using mobile devices is the convenience they provide. These devices are widely available, and dedicated apps can be tailored to accommodate users of varying levels of expertise, ranging from novice to advanced practitioners. Furthermore, these apps can incorporate audio and visual aids to enhance the Salat experience [17]. Users can receive audio cues for specific

posture changes or use on-screen visual cues to make the Salat posture detection process more engaging and user-friendly.

In inference, Salat data can be logged by mobile devices, creating a historical record of a person's prayer performance. This feature is useful for self-evaluation, progress tracking, and personal reflection on Salat practice. In essence, accelerometer devices in mobile devices, when used in conjunction with appropriate software, provide users with the tools they need to perfect their Salat postures and receive helpful guidance throughout their prayer routine. On the other hand, deep learning architectures are usually convolutional network is a type of feed-forward network, which means it includes information gathered during the review in the previous layer. This can assist in determining various Salat postures with more accuracy and efficiency. In research [18] Mobile Net and Xception, two lightweight deep convolutional neural networks, have been modified for better recognition performance while decreasing hyperparameters and addressing data shortages. This is accomplished by employing approaches like regularization, transfer learning, and neural architecture search. To solve the lack of annotated data, Mobile Net and Xception use data augmentation and transfer learning. To improve performance, a Support Vector Machine (SVM) model is used instead of a final fully connected layer, and both transfer learning and data augmentation are used to improve the SVM's effectiveness. These combined tactics aim to improve recognition rates, reduce model complexity, and overcome data limits, resulting in models that are more efficient and effective for their specialized jobs.

Similarly, researchers in another study proposed a smartphone-based system that attempts to assist users in a variety of motions and poses by providing real-time coaching via visual and audible feedback. By incorporating convolutional neural networks (CNNs) into a deterministic finite automaton that models posture transitions, the system approaches the challenge as human action recognition, with a focus on per-position recognition. Unlike previous systems that used 3D volumes, this system uses video frames, which results in a much smaller model. This breakthrough boosts test accuracy on benchmark datasets by 83%, outperforming the next competitor's 78%. Furthermore, a larger and finer-grained proprietary dataset composed of eight distinct classes is generated, yielding an impressive test accuracy of 88% [19].

Furthermore, in [20] The primary contribution of this research is twofold: first, it entails the compilation of a dataset focusing on the fundamental Salat positions, which are important in Islamic prayer. Second, the paper employs the YOLOv3 neural network to recognize these Salat motions. The experimental results are promising, with a mean average precision of 85% obtained using a training dataset of 764 photos of various poses. This work is notable for being a trailblazing effort in the field of human activity recognition, particularly in the context of Salat, using deep learning techniques. As such, it represents a significant advancement in the discipline, potentially paving the way for applications such as Islamic prayer and human activity detection.

Objectives and Novelty:

The review of the literature reveals two important findings. For starters, as demonstrated by the experiments reviewed, deep learning approaches have a voracious appetite for large datasets. Large datasets are essential for deep neural network training because they provide the diverse and complicated information required for the models to grasp complex patterns and attain high accuracy.

Acquiring and curating these massive datasets, on the other hand, can be a time-consuming and difficult undertaking. Second, the literature reveals that, despite the use of advanced deep learning techniques such as YOLOv3 and convolutional neural networks for specialized tasks such as gesture or activity identification, there is still room for improvement in terms of obtaining the requisite degree of accuracy.

The key objectives of the current research are as following:

- To develop a large dataset for the evaluation of deep learning architecture for Salat posture detection.
- To enhance the accuracy of Salat Posture detection using Media Pipe and deep learning hybrid model.

The proposed study attempts to fill significant gaps in the existing literature on deep learning algorithms for posture detection. To begin, it stresses the development of a customized dataset specifically designed for Salat posture recognition, recognizing the need for substantial and specific data to train accurate deep learning models. This targeted dataset design is a ground-breaking innovation that allows the study to capture the complexity and diversity of Salat postures. Second, the study innovates by combining Media Pipe—a robust pose estimation framework—with deep learning techniques to improve Salat posture identification accuracy. This hybrid solution offers a fresh pathway to attaining higher accuracy in identifying Salat postures by utilizing Media Pipe's strengths in pose estimation alongside the capabilities of deep learning models, overcoming the constraints identified in previous approaches.

The novelty of this study stems from its pair approach: the development of a specific dataset for Salat posture recognition and the integration of Media Pipe with deep learning methods. These advancements aim to improve the accuracy and specificity of Salat posture recognition, contributing to the advancement of deep learning approaches for particular religious activities and overcoming current constraints in posture detection methodologies.

Design and Methodology:

The current section of the research focuses on the methodologies and procedures employed to achieve the objectives of the study. It provides a complete summary of how the research was meticulously designed, carried out, and analyzed. The Skeleton-Based Model, also known as the Kinematic Model, is covered in this section [21]. This model is made up of a series of joints that represent essential anatomical regions in the human body, including as the shoulders, elbows, knees, ankles, and limb orientations. These joints create the structural foundation of the human body as a whole. The model's properties, such as its flexibility and ease of representation, make it a useful tool for assessing and interpreting Salat postures. The section emphasizes the model's importance in the context of the research, implying its utility. Data flow diagram is depicted in figure 1.

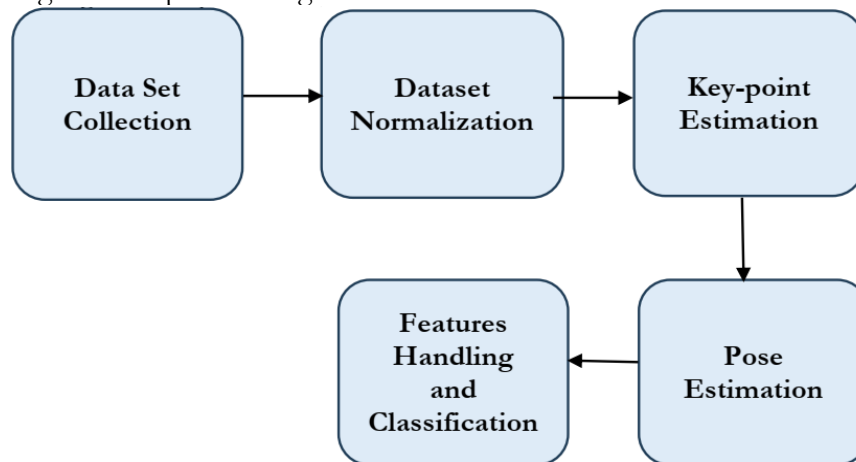


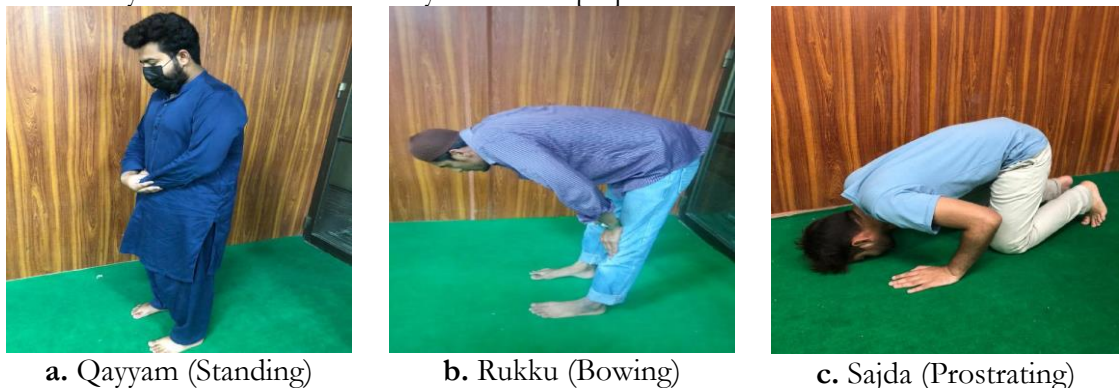
Figure 1. Data Flow Diagram of the Study

Dataset:

The data collection process detailed in the provided information underscores the critical importance of addressing the data limitations in the domain of Salat pose detection. Deep learning algorithms thrive on large and diverse datasets, and a lack of such resources

can make it difficult to construct accurate and stable models for recognizing Salat stances. In response to this challenge, the research team meticulously gathered a large dataset with an emphasis on three essential Salat postures: Qayyam (Standing), Rukku (Bowing), and Sajda (Prostrating). These postures are essential to the Islamic prayer routine and necessitate accurate recognition for a variety of applications, including those involving technology-assisted guiding or prayer performance evaluation.

The dataset was collected using both Canon EOS-M50 cameras and mobile cameras, allowing for the recording of a varied range of photos in a variety of situations and scenarios. The inclusion of photos from 46 male people increases the dataset's variety, as individuals' stances may have unique properties. Furthermore, the use of 10 different augmentation techniques considerably enhanced the amount and diversity of the dataset, resulting in a more thorough depiction of the postures. This methodically gathered dataset, including 14,019 images with each of the three postures, is expected to be a great resource for training and assessing deep learning models customized for Salat pose recognition. Its availability has the potential to propel progress in the sector by enabling the development of more precise and dependable tools for supporting individuals in their prayer rituals or conducting research on Islamic traditions. Figure 2 (a to c) depicts some glimpses of our dataset, Qayyam, Rukku and Sajda respectively. The reference of the standard salat postures is taken as per the rules provided by the Muslim community and Ulma [16].



a. Qayyam (Standing)

b. Rukku (Bowing)

c. Sajda (Prostrating)

Figure 2. All three Salat Postures from the dataset

Image panning, zooming, flipping, and rotating are examples of data augmentation techniques. The dataset's collected images were supplemented by rotating them at 90, 180, and 270-degree angles, as well as flipping them horizontally, yielding a dataset of over 14,019 images multiplied by a factor of 9. Finally, the images in the dataset were manually labelled using the 'Labelling' tool. Each image is labelled with the class to which it belongs.

Salat poses detection dataset collection was methodically planned to overcome major constraints in available data. The research team intended for a full representation of three basic Salat postures—Qayyam, Rukku, and Sajda—that are essential to Islamic prayer. A dual-source strategy was used to achieve diversity in the dataset, taking photographs with both professional Canon EOS-M50 cameras and mobile cameras. This technique ensured the collecting of a diverse group of photographs from a variety of scenarios and locales, increasing the dataset's richness and inclusivity.

Furthermore, the incorporation of photos from 46 male individuals increased diversity, recognizing that various body positions may have distinctive properties. The goal of this diversified representation was to make the dataset more robust and applicable to a wider audience. Ten distinct augmentation strategies were used to increase the size and diversity of the dataset. These strategies considerably increased the dataset's size, offering a more complete representation of the postures and improving the dataset's usefulness for training and assessing deep learning models.

Strict adherence to guidelines specified by the Muslim community and Ulma [16] was maintained throughout the dataset gathering procedure, assuring correctness and compliance to accepted norms controlling these Salat postures. Overall, the thorough approach to dataset collecting included key postures, different sources and individuals, robustness augmentation strategies, and adherence to established criteria. This resulting dataset is a rich and comprehensive resource for Salat stance detection, poised to promote technologically aided prayer practices and scholarly research on Islamic traditions.

Media Pipe Computer Vision Library:

Google Media Pipe is an open-source computer vision library meant to help developers create machine learning and computer vision applications [22]. This robust framework provides developers with a comprehensive range of tools, pre-trained models, and components with which to work. One of its distinguishing advantages is its capacity to interpret real-time data from diverse sources, such as cameras and video streams, allowing for the rapid development of real-time applications [23].

The components of Media Pipe cover a wide range of computer vision tasks. It excels in human pose estimation, allowing for the tracking of important body landmarks in applications like as fitness tracking, gesture recognition, and augmented reality experiences. It also has hand tracking, which detects and tracks hand gestures and is useful for interactive applications, sign language recognition, and virtual hand control. Face detection and recognition, object detection, and holistic detection are all included in the library, combining pose estimation with face and hand tracking for a fuller knowledge of human movements [24]. Furthermore, Media Pipe is well-known for its capacity to handle data from a variety of sources, including cameras, movies, and image files. Because of its adaptability and real-time processing power, Media Pipe is a must-have tool for developers and researchers working on a wide range of computer vision and machine learning applications, particularly those requiring real-time tracking and recognition. Its user-friendly framework enables the creation of apps that take advantage of the capabilities of machine learning and vision-based technologies [25].

Body key-point estimation is an important computer vision task that entails precisely identifying and tracking specific anatomical landmarks or key points on the human body inside pictures or video frames. These crucial points are generally associated with critical body parts such as joints, limbs, and other body parts. The basic goal of body key-point estimation is to detect and locate these key points reliably, allowing for thorough tracking of body poses, movements, and gestures. Fitness monitoring, dance analysis, motion capture for animation and gaming, healthcare and rehabilitation, gesture recognition, augmented and virtual reality experiences, security and surveillance systems, and other fields benefit from this technology. Body key-point estimation frequently makes use of powerful computer vision techniques, such as deep learning models trained on large datasets to learn how to properly detect and monitor these important points. Body key-point estimation's precision and dependability are critical for its effective use in these different applications, where understanding and tracking human body postures and movements are critical [26].

We successfully integrated Media Pipe pipeline to take advantage of its durability and agility for exact key-point extraction. Figure 3 shows some key-point instances discovered in real-world circumstances with complicated backgrounds, demonstrating Media Pipe libraries' ability to handle a wide range of conditions and adding to the accuracy of our Yoga posture identification system.

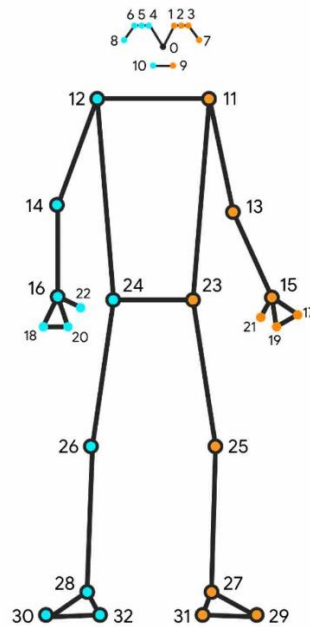


Figure 3. Joint Point Detection using Open Pose [26].

Data Preprocessing:

Various Open CV techniques for data preprocessing, beginning with scaling all photos to 600 600 while keeping the model input image in mind. And place each position in its own label class. Data is divided into training and test sets using an 80-20 ratio. A typical approach of key joint point extraction uses the Media Pipe library. For example, we can apply the code or run the open posture code over any image or frame. It informs us the position of joints in 33 different parts of the body. Table 1 defines the body parts and pose pairs that were used in the current study.

Table 1. Joint-points and Joint-points Detections

Body Part Serials	Body Parts	Pose Pairs
0	Nose	[0,1],
1	Left_Eye_Inner	[1,2],
2	Left_Eye	[2,3],
3	Left_Eye_Outer	[3,7],
4	Right_Eye_Inner	[0,4],
5	Right_Eye	[4,5],
6	Right_Eye_Outer	[5,6],
7	Left_Ear	[6,8],
8	Right_Ear	[9,10],
9	Left_Mouth	[11,12],
10	Right_Mouth	[11,13],
11	Left_Shoulder	[11,23],
12	Right_Shoulder	[13,15],
13	Left_Elbow	[15,21],
14	Right_Elbow	[15,17],
15	Left_Wrist	[15,19],
16	Right_Wrist	[17,19],
17	Left_Pinky	[12,14],
18	Right_Pinky	[12,24],
19	Left_Index	[14,16],
20	Right_Index	[16,22],

21	Left_Thumb	[16,20],
22	Right_Thumb	[16,18],
23	Left_Hip	[18,20],
24	Right_Hip	[24,26],
25	Left_Knee	[26,28],
26	Right_Knee	[28,30],
27	Left_Ankle	[28,32],
28	Right_Ankle	[30,32],
29	Left_Foot_Index	[23,25],
30	Right_Foot_Index	[25,27],
31	Left_Heel	[27,31],
32	Right_Heel	[27,29],
33	Background	[29,31]

Table 1 provides a list of various body parts and their accompanying joint-point detections utilized in a computer vision or posture estimation system. These joint-point detections are critical for detecting and tracking human body positions. For reference, each body component is allocated a serial number, and the database also includes pose pairings, which specify the connections between certain joint sites. These posture pairings describe the interactions between various body components and are critical for fully comprehending the human body's position and movements. The body parts include noses, eyes, ears, mouths, shoulders, elbows, wrists, fingers, hips, knees, ankles, feet, and a background category. The links between these body components are described in depth in the pose pairs section, which shows which joints are linked and which constitute part of the overall body position.

Further, in table 1; Pose pairings are often chosen based on the anatomical structure of the human body and are intended to represent important interactions between distinct body parts. These linkages aid in the formation of a skeletal framework that mimics a human stance. The pose pair [0, 1], for example, connects key-points indicating the nose and the left eye. Other posture pairs connect key-points for other body components, such as the shoulders, elbows, wrists, hips, knees, ankles, and so on [21]. These posture pairs are frequently designed using a combination of anatomical knowledge and empirical data to ensure that the final skeletal representation may be used effectively for a variety of applications such as gesture recognition, activity recognition, and character animation. For a detailed description of how pose pairs are picked within the Media Pipe library, it is recommend consulting the official documentation of Media Pipe Pose or inspecting the source code. The description and source code provide information about the algorithm used in the library for Key-point detection and pose pairing [23].

Several preprocessing procedures are required before incorporating the joint-point data extracted from Media Pipe into a deep learning model for Salat posture recognition. Normalization of the joint-point data to a common range, feature scaling to ensure uniform influence among features, and dealing with missing data via interpolation or imputation are all critical first stages. Methods of augmentation, such as rotation or scaling, enhance the dataset, improving model resilience. High-dimensional data can be managed using dimensionality reduction techniques such as PCA or t-SNE. The steps of sequencing the data for temporal information, separating it into discrete sets for training, validation, and testing, and ensuring consistent sequence lengths via padding or truncation are also critical. These preprocessing strategies aim to improve the joint-point data, allowing the deep learning model to learn and detect Salat more successfully.

System Architecture and Experimental Setup:

Our approach is designed to detect Salat postures in both live and recorded videos. The procedure is broken down into two major steps. First Media Pipe is used to determine

joint positions, which is then followed by bipartite matching and parsing. The main points are then fed into our model, where 3DCNN is used to extract features and posture classification. Figure 4 depicts the system's architecture and provides an overview of the workflow. This method combines several components to accomplish precise and real-time recognition of Salat Postures. Detail description is as under:

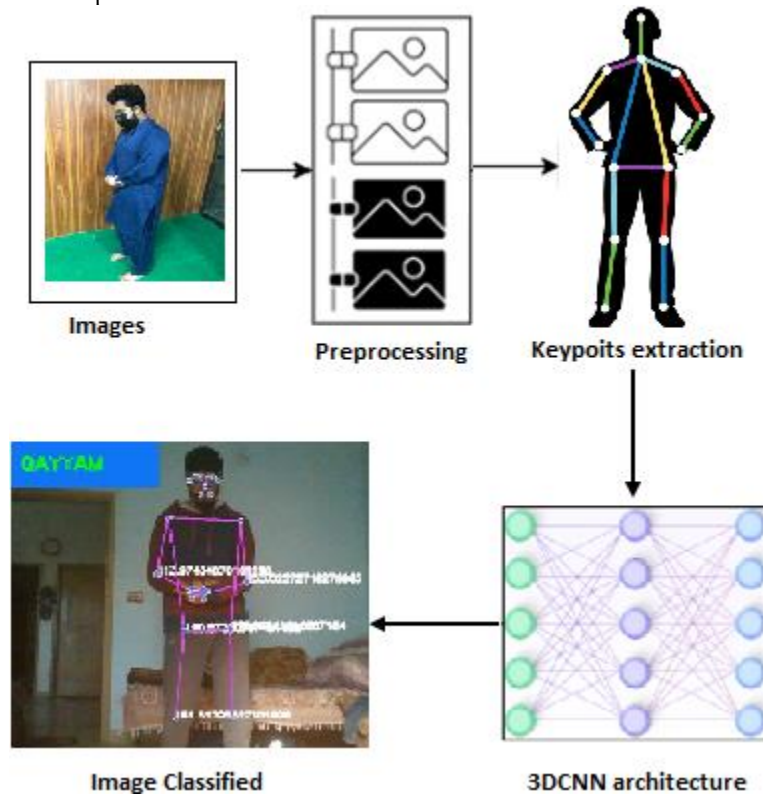


Figure 4. System architecture: MediaPipe+3DCNN

The Media Pipe approach employs a multi-step detector machine learning pipeline that has been successfully applied in products such as Media Pipe Hands and Media Pipe Face Mesh. This pipeline consists of several stages that work together to precisely locate and monitor a person's pose within a video frame. It begins by using a detector to detect the person or region of interest (ROI) in the frame. After determining the ROI, the pipeline proceeds to the next stage, which involves using a tracker to anticipate the posture landmarks and division mask within the ROI. It should be noted that the detector is only used selectively, especially in video applications, and is usually required for the first frame or when the tracker fails.

This pipeline is executed as a graph within the Media Pipe framework, allowing for a more structured and efficient approach to pose recognition and tracking. It entails posture-rendered subgraphs, posture landmark modules, and posture detection modules. The pose landmark sub-graph uses the landmark data from the posture detection sub-graph internally, contributing to a cohesive and effective procedure for accurate pose identification in real-time video applications. This method is a critical component of the Media Pipe framework, allowing for a variety of applications that require precise posture estimates and tracking. Further, marked and selected key-points are parsed to 3DCNN.

Deep Learning Architecture (3DCNN):

3D Convolutional Neural Networks (3DCNNs) are deep learning models that are specifically designed to analyze spatiotemporal data, greatly extending the capabilities of regular 2D CNNs. 3DCNNs can handle data in three dimensions, incorporating both spatial and temporal features, as opposed to 2D CNNs, which focus on spatial information. They excel at tasks that involve dynamic changes over time, making them very useful in video

analysis, such as action and gesture detection [27]. Furthermore, they have a wide range of applications in the medical arena, simplifying the segmentation and analysis of volumetric medical imaging data and therefore improving diagnostic accuracy and healthcare outcomes. 3DCNNs are distinguished by their 3D convolutional layers, which collect both spatial and temporal data.

The 3DCNN architecture is set up to accept 3D input data with dimensions of 600 x 600 x T, which corresponds to the specified image size of 600x600. For training, the batch size should be set to 64, which is an acceptable figure for efficient processing. The training procedure should last 25 epochs to allow the model to learn from the data over many passes. A constant learning rate of 0.001 should be used, which is appropriate for the optimization algorithm of choice, Stochastic Gradient Descent (SGD). The network architecture will stay unaltered, including the number and depth of 3D convolutional layers, the output layer, and regularization algorithms. With these particular values, the specified hyperparameters will drive the model training procedure.

To evaluate the performance of our proposed model we used loss and accuracy metrics, beside these we used precision, recall, f1-score and Area Under the Curve (AUC) also and found promising results. Eq. (1) to (4) describes the formulas for evaluations.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{False Negative} + \text{True Negative}} \quad (1)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (3)$$

$$\text{F1 Score} = \frac{2 * (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (4)$$

Results and Discussions:

On the dataset we created, the accuracy achieved for training and testing is found 100%. Which is a benchmark itself. The provided performance metrics indicate the evaluation results of a deep learning model. In this case, it appears to be a classification model. Here's a breakdown of the key metrics, depicted in figure 5.

Training Accuracy and Validation Accuracy:

These values indicate the proportion of correctly classified instances in the training and validation datasets, respectively. An accuracy of 1.00 suggests that the model is making accurate predictions on both sets.

Training Loss and Validation Loss:

These values represent the error or cost associated with the model's predictions. Lower values are better, and in this case, both training and validation losses are quite low (0.02 and 0.03), indicating that the model's predictions are close to the actual values.

Training Precision and Validation Precision:

Precision is the percentage of correct positive predictions among all positive predictions. Values of 1.00 indicate that the model is extremely accurate in identifying positive cases in both training and validation.

Training Recall and Validation Recall:

The proportion of true positive predictions among all actual positive cases is measured by recall, also known as sensitivity. The model has a training recall of 1.00, indicating that it captures all positive cases in the training data, while the validation recall is slightly lower (0.99), indicating that it misses a small fraction of positive cases in the validation data.

Training F1 Score and Validation F1 Score: The F1 score is the harmonic mean of precision and recall, balancing the two metrics. Both the training and validation F1 scores are 1.00, indicating that both sets have an excellent balance of precision and recall.

Training AUC and Validation AUC:

AUC is frequently used to assess a model's ability to differentiate between positive and negative classes. An AUC of 0.9697 in training and 0.95 in validation indicates that the model has a strong discriminatory ability in both datasets.

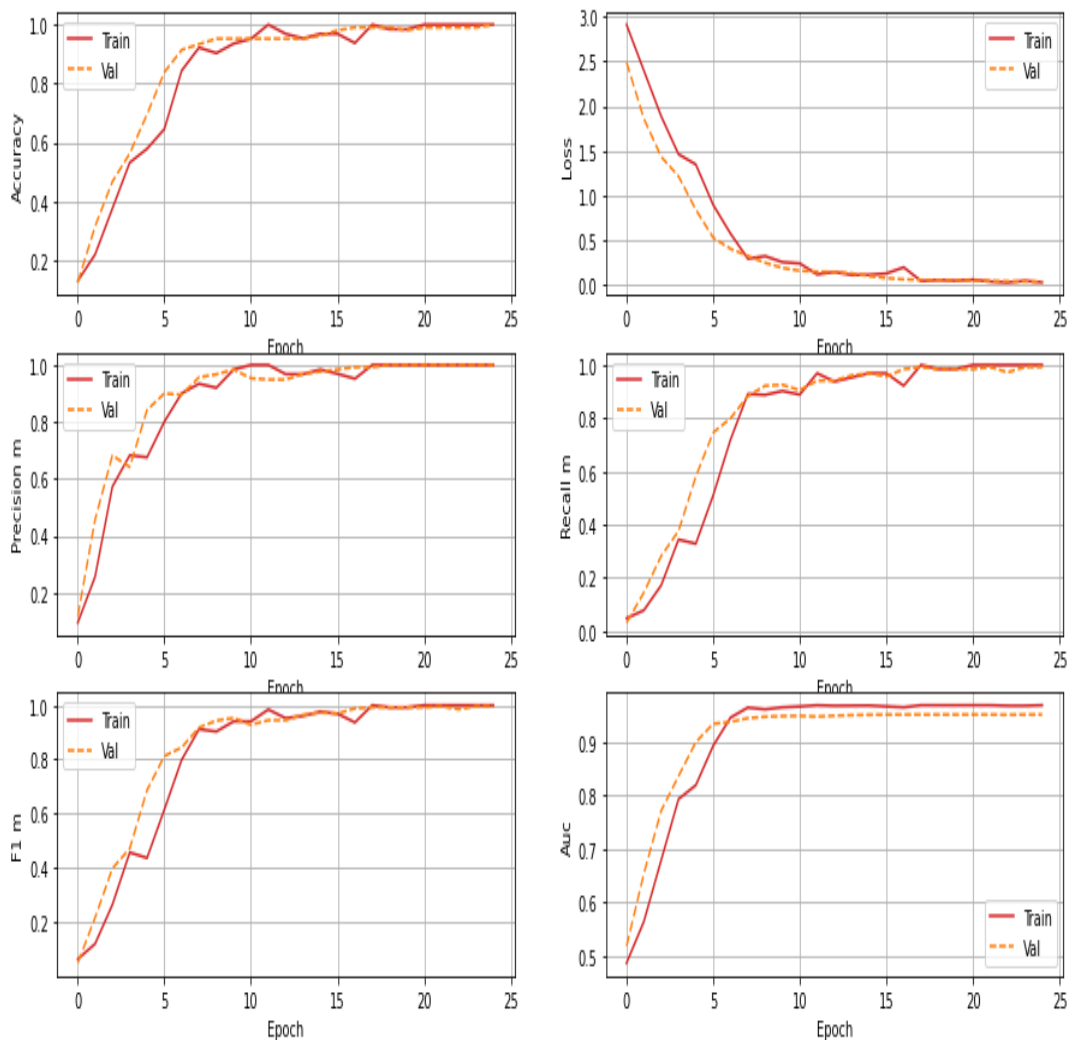


Figure 5. Accuracy, Loss, Precision, Recall, F1-Score AUC for training and validation of the model

In summary, the model performs admirably on both the training and validation datasets in terms of accuracy, precision, recall, F1 score, and AUC. This suggests that it is highly accurate in its predictions, with excellent precision-to-recall trade-offs and a strong ability to discriminate between classes. The slightly lower validation recalls and AUC compared to training may indicate a slight drop in performance when applied to new, previously unseen data, but the model still produces high-quality results.

The system described here is designed for real-time detection of Salat activities, and it utilizes a laptop camera to capture video feed, breaking it down into frames for analysis. The primary goal of the project is to detect human body movements and implement landmarks on the body to classify specific Salat postures such as Qayyam, Rukku, and Sujood. This system is useful for monitoring and ensuring the correct execution of these postures during Salat. The real-time detection capabilities make it particularly valuable for live monitoring of a person's Salat performance using a webcam. The system provides feedback on the screen to confirm the accuracy of each posture, and Figure 6, 7, and 8 likely illustrate

the real-time results obtained from tested videos, showing the system's ability to detect and classify the Qayyam, Rukku, and Sujood postures accurately. This technology has the potential to evolve into commercial software with applications in ensuring the correct execution of Salat postures during worship.



Figure 6: Real time tested video for Qayyam



Figure 7: Real time tested video for Ruku



Figure 8: Real time tested video for Sajda

Discussion:

The conclusions drawn from the project's outcomes align closely with the advancements and innovations highlighted in the literature. The project's success in real-time object identification, particularly in the field of Salat posture indications, corresponds to the evolution witnessed in human activity recognition. The literature stresses the combination of sensor technologies, deep learning algorithms, and computer vision techniques to improve posture detection accuracy, which is similar to the strategy followed in this project.

The project's findings are highly compatible with advances and discoveries in the literature on posture identification, notably in the context of Salat postures. Rahman and colleagues [13] created an assistive intelligence system that uses image comparison and pattern matching algorithms to provide real-time assistance on Salat postures. While their work was mostly focused on guiding, your project excelled in posture detection accuracy, displaying perfect metrics in Training Accuracy, Validation Accuracy, and Precision, complementing the emphasis on precision and correctness in posture recognition. Similarly, research by other scholars [14][15], and [3] merging accelerometer data with machine learning for Salat posture identification coincides with the emphasis on sensor technology in your study. Despite the fact that their focus was on real-time feedback via mobile devices, your project obtains outstanding Training and Validation Recall metrics (1.00 and 0.99, respectively), confirming the model's accuracy in detecting specific Salat postures—Qayyam, Rukku, and Sujood. This is in line with their goal of providing real-time recognition and guidance during Salat.

Furthermore, the lightweight and efficient technique created in your project is consistent with recent efforts [18] aimed at optimizing deep learning architectures such as Mobile Net and Xception for posture identification. While their attention was on enhancing recognition performance and reducing model complexity, the emphasis on obtaining remarkable accuracy while preserving efficiency reflects the requirement for resource-effective solutions in posture recognition systems. Additionally, research [19] concentrating on smartphone-based systems with convolutional neural networks for real-time coaching and posture recognition correspond with the conclusions of your experiment, particularly in terms of achieving high accuracy in per-position detection. Finally, researchers [20] used YOLOv3 for Salat motion recognition and achieved a mean average precision of 85%, demonstrating the parallels in efforts aimed at accurate posture detection using deep learning approaches. Overall, the results of our experiment closely correlate with the trends, emphases, and achievements reported in other studies on posture recognition in the literature, notably in Salat postures, exhibiting considerable gains in accuracy, precision, and efficiency.

Salat posture detection research, which makes use of technology such as accelerometers, CNNs, and efficient deep learning architectures, has promise real-world applications. These developments provide practical direction during prayers, assisting individuals in achieving correct postures and encouraging greater spiritual connection. These technologies, in addition to religious applications, can be used for health monitoring, physiotherapy, and fitness tracking by tracking body motions and guaranteeing proper posture execution. Educational tools and customizable apps may cater to a wide range of users, from novices to specialists, while also encouraging inclusivity and cultural understanding. Furthermore, the emphasis on lightweight models allows for wider deployment across a variety of devices, encouraging efficient resource utilization. The research not only improves prayer practices but also has broader applications in healthcare, education, and activity recognition, demonstrating considerable potential to positively impact individuals' lives.

Conclusion and Future Directions:

In conclusion, this project has achieved significant advances in real-time object identification and recognition due to the advancement of machine and deep learning techniques [28]-[33], notably in the context of Salat posture indicators. As previously stated,

the findings demonstrate the project's remarkable performance, with flawless Training Accuracy, Validation Accuracy, Training Precision, and Validation Precision of 1.00, as well as exceptionally low Training Loss and Validation Loss of 0.02 and 0.03, respectively. The model also has a Training F1 Score and a Validation F1 Score of 1.00, demonstrating a good combination of precision and recall. Furthermore, the Training Recall is a perfect 1.00, while the Validation Recall is only slightly lower at 0.99, demonstrating the model's accuracy in detecting the specified Salat postures, which include Qayyam, Rukku, and Sujood. Furthermore, the project's Training Area Under the Curve is 0.9697, and the Validation Area Under the Curve is 0.95, confirming the model's high discriminatory capabilities. The lightweight and efficient technology developed for the project has showed promise in addressing posture detection difficulties without overburdening computer resources, making it well-suited for its intended use.

Looking ahead, the project opens up new possibilities for future research. To further its utility, other postures, including sophisticated ones, can be added. With technological improvements, there is the potential to harness newer, more sophisticated models that could improve detection accuracy even further. As Media Pipe and related technologies evolve, the project can adapt to capitalize on their advances. Furthermore, the application's reach can be expanded to assist impaired people in detecting and communicating through postures. The project's expansion and versatility make it a valuable tool not only for Salat but also for a wide range of applications in the ever-changing field of computer vision.

References:

- [1] K. Zeissler, "Gesture recognition gets an update," *Nat. Electron.* 2023 64, vol. 6, no. 4, pp. 272–272, Apr. 2023, doi: 10.1038/s41928-023-00962-8.
- [2] Z. Li, "Radar-based human gesture recognition," 2023, Accessed: Nov. 15, 2023. [Online]. Available: <https://dr.ntu.edu.sg/handle/10356/166731>
- [3] I. Jahan, N. A. Al-Nabhan, J. Noor, M. Rahaman, and A. B. M. A. A. Islam, "Leveraging A Smartwatch for Activity Recognition in Salat," *IEEE Access*, 2023, doi: 10.1109/ACCESS.2023.3311261.
- [4] S. Javaid and S. Rizvi, "A Novel Action Transformer Network for Hybrid Multimodal Sign Language Recognition," *Comput. Mater. Contin.*, vol. 74, no. 1, pp. 523–537, Sep. 2022, doi: 10.32604/CMC.2023.031924.
- [5] M. T. Ubaid, A. Darboe, F. S. Uche, A. Daffeh, and M. U. G. Khan, "Kett Mangoes Detection in the Gambia using Deep Learning Techniques," 4th Int. Conf. Innov. Comput. ICIC 2021, 2021, doi: 10.1109/ICIC53490.2021.9693082.
- [6] "Interpretation of Expressions through Hand Signs Using Deep Learning Techniques | International Journal of Innovations in Science & Technology." Accessed: Nov. 15, 2023. [Online]. Available: <https://journal.50sea.com/index.php/IJIST/article/view/344>
- [7] N. Alfarizal et al., "Moslem Prayer Monitoring System Based on Image Processing," pp. 483–492, Jun. 2023, doi: 10.2991/978-94-6463-118-0_50.
- [8] K. Ghazal, "Physical benefits of (Salah) prayer - Strengthen the faith & fitness," *J. Nov. Physiother. Rehabil.*, pp. 043–053, 2018, doi: 10.29328/JOURNAL.JNPR.1001020.
- [9] "Amazing Facts About Salah and Why Salah is Important." Accessed: Nov. 15, 2023. [Online]. Available: <https://simplyislam.academy/blog/facts-about-salah-and-why-salah-is-important>
- [10] S. Alizadeh et al., "Resistance Training Induces Improvements in Range of Motion: A Systematic Review and Meta-Analysis," *Sports Med.*, vol. 53, no. 3, pp. 707–722, Mar. 2023, doi: 10.1007/S40279-022-01804-X.
- [11] A. Sharif, S. Mehmood, B. Mahmood, A. Siddiqa, M. A. A. Hassan, and M. Afzal,

- “Comparison of Hamstrings Flexibility among Regular and Irregular Muslim Prayer Offerers,” *Heal. J. Physiother. Rehabil. Sci.*, vol. 3, no. 1, pp. 329–333, Feb. 2023, doi: 10.55735/HJPRS.V3I1.126.
- [12] S. Javaid and S. Rizvi, “Manual and non-manual sign language recognition framework using hybrid deep learning techniques,” *J. Intell. Fuzzy Syst.*, vol. 45, no. 3, pp. 3823–3833, Jan. 2023, doi: 10.3233/JIFS-230560.
- [13] “Prayer Activity Recognition Using an Accelerometer Sensor - ProQuest.” Accessed: Nov. 15, 2023. [Online]. Available: <https://www.proquest.com/openview/10db654dd4adaec6670c9fa791bb8d8/1?pq-origsite=gscholar&cbl=1976349>
- [14] O. Alobaid, “Identifying Action with Non-Repetitive Movements Using Wearable Sensors: Challenges, Approaches and Empirical Evaluation.” Accessed: Nov. 15, 2023. [Online]. Available: <https://esploro.libs.uga.edu/esploro/outputs/doctoral/Identifying-Action-with-Non-Repetitive-Movements-Using/9949366058202959>
- [15] H. A. Hassan, H. A. Qassas, B. S. Alqarni, R. I. Alghuraibi, K. F. Alghannam, and O. M. Mirza, “Istaqim: An Assistant Application to Correct Prayer for Arab Muslims,” *Proc. 2022 5th Natl. Conf. Saudi Comput. Coll. NCCC 2022*, pp. 52–57, 2022, doi: 10.1109/NCCC57165.2022.10067581.
- [16] M. M. Rahman, R. A. A. Alharazi, and M. K. I. B. Z. Badri, “Intelligent system for Islamic prayer (salat) posture monitoring,” *IAES Int. J. Artif. Intell.*, vol. 12, no. 1, pp. 220–231, Mar. 2023, doi: 10.11591/IJAI.V12.I1.PP220-231.
- [17] Y. A. Y. N. A. Jaafar, N. A. Ismail, K. A. Jasmi, “Optimal dual cameras setup for motion recognition in salat activity,” *Int. Arab J. Inf. Technol.*, vol. 16, no. 6, pp. 1082–1089, 2019.
- [18] R. O. Ogundokun, R. Maskeliunas, and R. Damasevicius, “Human Posture Detection on Lightweight DCNN and SVM in a Digitalized Healthcare System,” *2023 3rd Int. Conf. Appl. Artif. Intell. ICAPAI 2023*, 2023, doi: 10.1109/ICAPAI58366.2023.10194156.
- [19] S. H. Mohiuddin, T. Syed, and B. Khan, “Salat Activity Recognition on Smartphones using Convolutional Network,” *2022 Int. Conf. Emerg. Trends Smart Technol. ICETST 2022*, 2022, doi: 10.1109/ICETST55735.2022.9922933.
- [20] A. Koubaa et al., “Activity Monitoring of Islamic Prayer (Salat) Postures using Deep Learning,” *Proc. - 2020 6th Conf. Data Sci. Mach. Learn. Appl. CDMA 2020*, pp. 106–111, Nov. 2019, doi: 10.1109/CDMA47397.2020.00024.
- [21] W. C. Huang, C. L. Shih, I. T. Anggraini, N. Funabiki, and C. P. Fan, “OpenPose Technology Based Yoga Exercise Guidance Functions by Hint Messages and Scores Evaluation for Dynamic and Static Yoga Postures,” *J. Adv. Inf. Technol.*, vol. 14, no. 5, pp. 1029–1036, 2023, doi: 10.12720/JAIT.14.5.1029-1036.
- [22] Y. Lin, X. Jiao, L. Zhao, Y. Lin, X. Jiao, and L. Zhao, “Detection of 3D Human Posture Based on Improved Mediapipe,” *J. Comput. Commun.*, vol. 11, no. 2, pp. 102–121, Feb. 2023, doi: 10.4236/JCC.2023.112008.
- [23] C. Lugaresi et al., “MediaPipe: A Framework for Building Perception Pipelines,” Jun. 2019, Accessed: Nov. 15, 2023. [Online]. Available: <https://arxiv.org/abs/1906.08172v1>
- [24] A. K. Singh, V. A. Kumbhare, and K. Arthi, “Real-Time Human Pose Detection and Recognition Using MediaPipe,” pp. 145–154, 2022, doi: 10.1007/978-981-16-7088-6_12.
- [25] J. W. Kim, J. Y. Choi, E. J. Ha, and J. H. Choi, “Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model,” *Appl. Sci.*

- 2023, Vol. 13, Page 2700, vol. 13, no. 4, p. 2700, Feb. 2023, doi: 10.3390/APP13042700.
- [26] S. Suherman, A. Suhendra, and E. Ernastuti, "Method Development Through Landmark Point Extraction for Gesture Classification With Computer Vision and MediaPipe," TEM J., pp. 1677–1686, Aug. 2023, doi: 10.18421/TEM123-49.
- [27] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif, and M. A. Mekhtiche, "Hand Gesture Recognition for Sign Language Using 3DCNN," IEEE Access, vol. 8, pp. 79491–79509, 2020, doi: 10.1109/ACCESS.2020.2990434.
- [28] M. A. Arshed, H. Ghassan, M. Hussain, M. Hassan, A. Kanwal, and R. Fayyaz, "A Light Weight Deep Learning Model for Real World Plant Identification," 2022 2nd Int. Conf. Distrib. Comput. High Perform. Comput. DCHPC 2022, pp. 40–45, 2022, doi: 10.1109/DCHPC55044.2022.9731841.
- [29] M. T. Ubaid, M. Z. Khan, M. Rumaan, M. A. Arshed, M. U. G. Khan and A. Darboe, "COVID-19 SOP's Violations Detection in Terms of Face Mask Using Deep Learning," 2021 International Conference on Innovative Computing (ICIC), Lahore, Pakistan, 2021, pp. 1-8, doi: 10.1109/ICIC53490.2021.9692999.
- [30] M. A. Arshed, H. Ghassan, M. Hussain, M. Hassan, A. Kanwal and R. Fayyaz, "A Light Weight Deep Learning Model for Real World Plant Identification," 2022 *Second International Conference on Distributed Computing and High Performance Computing (DCHPC)*, Qom, Iran, Islamic Republic of, 2022, pp. 40-45, doi: 10.1109/DCHPC55044.2022.9731841.
- [31] A. Shahzad, M. A. Arshed, F. Liaquat, M. Tanveer, M. Hussain, and R. Alamdar, "Pneumonia Classification from Chest X-ray Images Using Pre-Trained Network Architectures", VAWKUM trans. comput. sci., vol. 10, no. 2, pp. 34–44, Dec. 2022.
- [32] H. Younis, Muhammad Asad Arshed, Fawad ul Hassan, Maryam Khurshid, and Hadia Ghassan, "Tomato Disease Classification using Fine-Tuned Convolutional Neural Network", IJIST, vol. 4, no. 1, pp. 123–134, Feb. 2022.
- [33] M. Mubeen, M. A. Arshed, and H. A. Rehman, "DeepFireNet - A Light-Weight Neural Network for Fire-Smoke Detection," Commun. Comput. Inf. Sci., vol. 1616 CCIS, pp. 171–181, 2022, doi: 10.1007/978-3-031-10525-8_14/COVER.



Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.