Research Design 2

Ryan Galea
Institute of Information & Communication Technology
Malta College or Arts, Science & Technology
Corradino Hill
Paola PLA 9032
ryan.galea.c10264@mcast.edu.mt

***Abstract*—The rising prevalence of overweight and obesity in Malta, coupled with diminishing physical activity levels, signifies a public health concern. This study introduces a machine learning application to combat this issue by promoting home-based static exercises. We leverage pose estimation technology to ensure the safety and effectiveness of these exercises, mitigating the risk of injuries from improper postures. This innovative solution provides a flexible and safe fitness option, eliminating the need for professional supervision, and enhancing individuals' propensity to engage in regular physical activity.**

***Index Terms*—Image Processing, Computer Vision, Machine Learning, MediaPipe**

I. Introduction

This conference paper centers around the use of machine learning for correcting exercise posture, specifically the plank exercise at home. The grounds behind this topic is the dire need for safer, effective, and accessible fitness methods during a time where home workouts have become increasingly popular.

Guiding the methodology of this study, the deductive approach focuses on testing the hypothesis that a machine learning software can correct plank postures effectively. Selecting a case study as the strategy, the developed software prototype forms the core of the investigation. Furthermore, the mono method leverages qualitative data from technical resultant data to understand performance and efficiencies in depth. A cross-sectional time was chosen due its ability to snapshot computational technology and software efficacy at a specific point in time. The techniques and procedures in this research involve analysing result data, testing the software prototype, and then undertaking a detailed analysis result of the collected qualitative data. This strategically chosen positioning within the Research Onion ensures a detailed exploration in an effective manner.

The background of the research theme relies on several factors that mainly inspired from the global health crisis in which quarantine has led to lack of physical activities in daily routines. The high rates of obesity in Malta were also an inspiration to allow more flexible exercises that involves less commitment. These pressing health concerns have fueled the necessity for innovative fitness solutions that are both accessible and safe for users. Furthermore, the reason that static exercises were chosen was due to it being accessible and effective to most users and an ideal as an introductory exercise to develop.

The objective of this study was to evaluate whether a machine learning-based software solution could evaluate in real-time plank stance corrections, hence enabling safe, easily accessible, and more effective home workouts.

When addressing the purpose and aim of this research, it would classify as twofold. Firstly, to develop and test a prototype that could potentially revolutionize home workout practices by providing real-time feedback on exercise posture. Secondly, to investigate the user experience and effectiveness of this tool to validate the research hypothesis and contribute to individual fitness and well-being at home.

This conference paper will delve into this research, dissecting its findings, and providing a critical evaluation in comparison to similar studies.

II. Literature Review

The field of pose estimation has gained popularity amongst AI methodologies, each with its unique approach to the problem. The exploration of these methodologies provides a broader context for the work undertaken by my paper and highlights a good level of diversity among the techniques in this research area.

Narayanan et al. [1] decided to tackle the challenging issue of incorrect posture while exercising. They used the latest tech in computer vision and deep learning to create a unique solution. Their focus was yoga, a beneficial physical activity which can be potentially harmful if performed incorrectly.

In their study, OpenPose was used to spot and identify different human body parts. OpenPose was originally designed for sports and security uses but the team adapted it for yoga pose detection. The software breaks down the human form into a bunch of key points and generates a skeleton-like image of the body.

This system uses Confidence Maps and Part Affinity Fields to refine its predictions. Confidence Maps are a visual representation of where the system thinks a particular body part is, while Part Affinity Fields are used to encode the position and direction of limbs.

Go and Aoki [2] focused their research on a methodology for human pose estimation that is composed of several techniques. These include, Convolutional Neural Network (CNN) based pose estimation, Bounding-Box Curriculum Learning (BCL), and Recurrent Pose Estimation (RPE).

A CNN architecture called DeepPose is utilized for top-view human pose estimation. It consists of 5 convolutional layers and 2 fully connected layers. The model estimates the coordinates of 14 human joints from top-view images, producing a 28-dimensional vector as output.

To mitigate instability in the model's performance due to challenging human poses, RPE is employed. This technique is based on the hypothesis that there is a correlation between the rectangle of predicted joints (predicted box) and the bounding-boxes. Depending on the ratio of the sides of the predicted box and the bounding-box, RPE adjusts the bounding-box size by either extension or reduction. This process continues until the estimator finds the proper size for estimation.

Liu et al. [3] created an automatic pipeline that works directly from raw video sequences rather than being limited to pre-selected human-contained bounding boxes. This innovative design encompasses various preprocessing steps, moving away from most existing works in the field.

The system copes with lighting variation problems during the day and night, and this was addressed through the introduction of the Infrared Selective (IRS) acquisition method which helps to ensure stable image quality under varying illumination conditions, thereby facilitating reliable and robust performance estimation.

To manage unusual pose distributions from overview angles, they proposed the n-end Histogram of Oriented Gradients (HOG) rectification method, which effectively rectifies the orientation of in-bed images. This process facilitates the use of existing pose estimation models without the need for extensive training, thus increasing the efficiency of the pipeline.

They repurposed a general-purpose pose estimation model based on deep neural networks for in-bed pose estimation, utilising a pretrained Convolutional Pose Machine (CPM) and a high-performance pictorial structure-oriented method known as Flexible Mixture of Parts (FMP). These two algorithms represent two typical frameworks for pose estimation, one based on deep learning CPM and the other on the pictorial structure model FMP.

In their study Ukita et al. [4] propose a method to enhance the efficiency of stochastic models used for recognition. They aim to reduce the computational cost of training these models by carefully selecting and decreasing the number of training samples. Traditionally, training these models would require a large number of samples, which results in significant computational costs and time expenditure. For instance, training a model with a dataset like ImageNet, which consists of 1.2 million images across 1000 object classes, could take upwards of 250 days with traditional image features and classifiers.

The authors argue that reducing the number of training samples shouldn't be done randomly, as that can adversely affect the recognition accuracy. Instead, they suggest selecting samples that are effective for discriminating between recognition classes. The hypothesis is that even with a small number of such samples, a model can still perform accurate recognition.

A major consideration in the authors' proposed method is the trade-off between computational cost and recognition accuracy. In trials for model selection, they contend that absolute accuracy is not necessarily required, but the relative accuracy of different models should be verified.

Their method introduces the concept of 'learning potential models' with selected samples, which is significantly faster than learning with all samples. The best model, which would have the highest accuracy in trials, can then be retrained with all samples. Although this would involve training that model twice, the overall learning time is reduced as learning with selected samples is much quicker than learning with all samples. Consequently, the total cost with selected samples is lower than with all samples.

The proposed method is applied to human pose estimation in still images due to its wide range of applications and complexity. Human pose estimation, which requires modelling a large number of parameters, can benefit significantly from efficient modelling methods like the one proposed by the authors.

However the iterative nature of this learning method must be trained one by one with the latent SVM for semi-convex optimization, leading to a substantial computational cost. By reducing the number of training samples and following their proposed method of sample selection, the authors aim to significantly reduce this computational cost, achieving efficient learning and high accuracy in human pose estimation.

Munea et al.'s [5] discusses a comprehensive exploration of various methodologies employed in 2D human pose estimation. Deep learning has played a significant role in the advancements of human pose estimation. For instance, the initial exploration with DeepPose leveraged the AlexNet architecture. Further research took a leap and explored the realms of R-CNN, Fast R-CNN, FPN, Faster R-CNN, and Mask R-CNN. A particular emphasis was laid upon VGG and ResNet architectures, which have become prevalent in recent studies due to their performance efficiency and robustness.

The authors' study also delves into the two primary strategies for pose estimation which primarily isolates individuals in an image using an object detector before focusing on computing their poses.

Several innovative models are discussed within the body of their work. Among these, the ConvNet architecture stands out for its unique approach in generating discrete heatmaps for each joint location in a given image. Another approach, CPM, creates a 2D belief map for each keypoint's location, leveraging a sequence of convolutional networks. It is designed to learn long-range spatial relationships, addressing the challenge of vanishing gradients with intermediate supervision after each stage. Meanwhile, the Stacked Hourglass network takes advantage of varying scales to capture a multitude of information from an image.

In their study, Dawanage et al. [6] look into into the process of human pose estimation, a field of study gaining significant attention in the machine learning community. The team utilizes several established techniques and the latest methodologies to recognize human poses from both images and videos.

They approach this by first analysing various prior studies, starting from 2017. These investigations highlight the use of deep convolutional neural networks, which have evolved to become the dominant solution in the field. From the established methods, the team identifies two principal strategies, regression of key-point positions and estimation of key-point heatmaps.

In their methodology, they chose to use BlazePose, a recent model developed by Google that functions seamlessly on lightweight devices such as browsers or mobile devices.

BlazePose, described as a real-time pose detection technique, works by detecting specific body parts like elbows, hips, wrists, knees, ankles, and subsequently forming a skeletal structure of the pose by connecting these points. BlazePose uses depthwise separable convolution to deepen the network and reduce parameters, hence lowering computational cost and boosting accuracy.

With the use of MediaPipe framework that known for building multimodal machine learning pipelines, they were able to provide the the basis for implementing state-of-the-art models, including human face detection, multi-hand tracking, hair segmentation, object detection, and tracking. The combination of BlazePose and MediaPipe is at the heart of the methodology presented in their study.

III. Methodology

In light of increasing sedentary lifestyles and an obesity pandemic, this paper delves into three core research questions:

1) How does the camera angle affect body detection and recognition accuracy within an image-processing exercise aid?
2) Can a software system feasibly replace a personal trainer in supervising the execution of plank exercises?
3) How does the recognition system perform under various lighting conditions, specifically in excessively lit or dim environments?

To examine and choose an appropriate model from existing pose estimation algorithms. To analyze current datasets and libraries, selecting the most suited options for real-time body recognition and feedback. To identify and prioritize recognition methods that can deliver real-time detections and instant feedback. In the realm of research philosophies, this study is positioned within the positivism paradigm, recognizing that research is a problem-solving activity. The positivism approach emphasizes the application of what works best in practice, allowing the combination of different methods to understand the research problem from multiple dimensions. This resonates with this study, as it seeks to deliver a practical, technological solution to a health concern.

This study employs a methodology with mixed-methods approach, combining quantitative data that's derived from the angles and timing of performed exercises, with qualitative data, sourced from the assessment of correct and incorrect postures.

Python forms the backbone of this investigation, with the OpenCV library providing the tools for computer vision and body tracing. The implementation further leverages the power of MediaPipe, a machine learning framework developed by Google. This pre-trained model obviates the need for self-training, saving considerable time and effort, and offers a robust dataset for body tracing.

The software captures a video feed from a laptop's webcam, which is then adjusted to support RGB colors. Body landmarks are then extracted by the MediaPipe framework to visually depict pose tracking. The model overlays the identified body areas with a skeleton-like diagram, providing a visual interpretation of the recognition process.

With reference of answering the research questions, the applied method supported RGB colors, making image differentiation better. However, there may be some limitations when it comes to other exercises and possible mistakes in detecting body shapes.

This study also proved that the system can give real-time feedback on how exercises are performed. It could replace a personal trainer, but more research is required to see how it works with different exercises and people.

Lastly the study did test the system under various lighting conditions. Specifically, the system was assessed using participant 2 in a low-lit environment. Still, more tests across different light situations would make the system even stronger and more reliable for all sorts of real-world conditions.

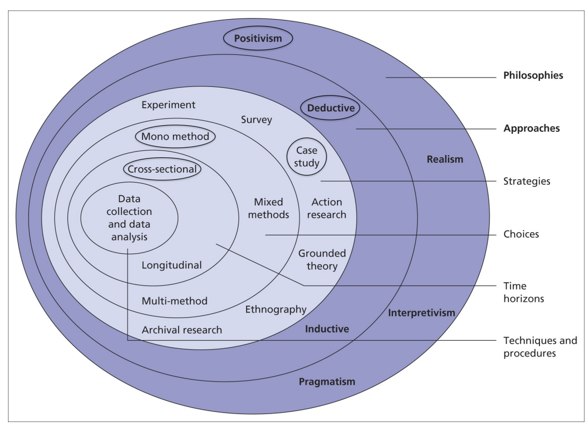

Fig. 1. Research Onion

This study's ensures proper ethical considerations by ensuring integrity and trustworthiness by keeping in line with the General Data Protection Regulation (GDPR) as per the European Union law. Due the possible sensitivity of the video footage, data security is of the utmost importance. This study reduces the 'black box' worry frequently associated with AI and machine learning systems since nothing is stored while analysing in real-time. The input data is only used for training and safely constricted from unauthorised access with utmost confidentiality.

The method minimises any potential biases by accommodating a wide range of body types and postures, by not discriminating based on factors such as gender, size, shape, or any other aspect of an individual. This was well taken into consideration during the design process because the software's main goal is to promote physical health by supporting workout postures.

Moreover, informed permission is essential, being aware of the potentially sensitive nature of the video material used, participants' explicit agreement was obtained for the study's purposes. Careful precautions were taken to protect personal information throughout data collection, processing and visualisation, assuring identity protection.

Last but not least, it's critical to stress that the information presented regarding the software's potential and restrictions is accurate and truthful, avoiding misinformation about its capabilities to proudly comply ethical principles.

IV. Results

In assessing the final prototype of the software, two individuals with differing lifestyles and levels of physical activity were interviewed. Participant 1, a 26-year-old with a more sedentary lifestyle, and Participant 2, a 19-year-old active in sports and training, provided in-depth qualitative responses that offered more detailed insights than quantitative data might have.

Concerning exercise habits and perceptions of public health, Participant 1 and Participant 2 presented differing opinions on the prevalence of exercise, while both agreed on the inadequacy of government initiatives to promote physical activity. Participant 2 suggested incentives such as reduced gym membership prices and lower tax rates on healthy foods. When discussing the obesity issue in Malta, both individuals promptly confirmed its presence, corroborating a 2016 study on the prevalence of obesity in Malta, especially among men.

Participant 1 and Participant 2's exercise preferences offered insight into potential use cases for the software. Participant 1, preferring home workouts but lacking confidence to train alone, embodies a significant target demographic for this software. Participant 2, although utilizing a gym for specific equipment, showed openness to a hybrid model of home and gym workouts, should equipment not be a factor. These responses suggest an overall preference for home workouts, particularly for those not sufficiently confident or informed to train alone, thereby reinforcing the need for a solution like the proposed software.

Speaking to experiences of injury, Participant 2 recounted an incident caused by incorrect machinery posture, reinforcing the earlier referenced study's findings about the dangers of improper exercise forms. This highlights the potential for the software to help users avoid injury.

Both interviewees responded positively to the prospect of using a phone-based exercise assistance tool. Though their exercise preferences varied, with Participant 1 favoring static exercises and Participant 2 more open to varied workout routines, both saw value in the proposed application's capabilities.

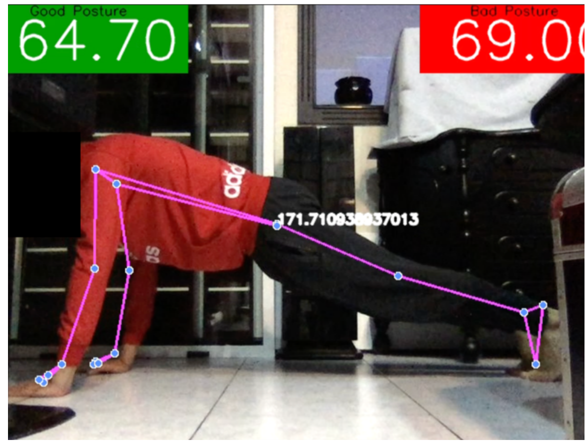

Fig. 2. Participant 1 in an correct Plank Posture

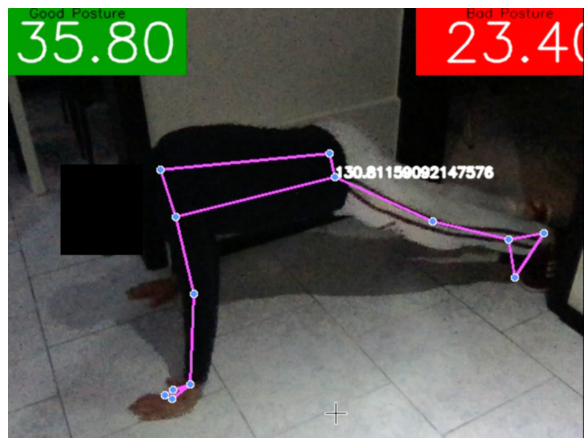

Fig. 3. Participant 2 in an incorrect Plank Posture

The results of this research clearly demonstrate the effectiveness of the MediaPipe framework for human pose estimation tasks. Comparatively, the surveyed methodologies in existing literature employ varied and inventive methods to solve similar problems, offering unique insights that reinforce the advantages of the approach adopted in this research.

In comparison, Narayanan et al. [1] showcased an innovative application of OpenPose for yoga posture detection and correction. Their system is interesting for its real-time functionality, a trait shared with this study's framework. However, the newly improved efficiency and accuracy provided by MediaPipe places the approach of this study at an advantage.

Similarly, Go and Aoki [2] introduced the combination of CNN-based pose estimation, BCL, and RPE for human pose estimation. Their methodology shares commonalities with this study, specifically the CNN-based approach. However, the intricate bounding-box strategies and recurrent pose adjustments they employed can be computationally intense. In comparison, the MediaPipe framework chosen in this study capitalizes on efficiency without compromising on accuracy.

The work of Liu et al. [3] brought an inventive perspective to the field by developing an automatic pipeline that can work directly from raw video sequences, tackling problems like unusual pose distributions and varying lighting conditions. Their focus on in-bed pose estimation highlighted the specific difficulties of this domain, which might not be applicable in broader contexts, as in our research. Nonetheless, the techniques they used serve as a reminder of the importance of versatility in pose estimation methodologies.

In focus to reduce computational costs, Ukita et al. [4] proposed a careful selection and decrease in the number of training samples. While their method offers a way to manage the trade-off between computational cost and recognition accuracy, it may lead to an undersampling problem, potentially reducing the robustness of the trained model. By contrast, in this research did not require such a trade-off, thanks to the efficiency of MediaPipe.

Munea et al.'s [5] detailed exploration of various methodologies and backbone architectures in 2D human pose estimation further emphasizes the importance of deep learning and the selection of a suitable model. Their in-depth examination of different models and the uniqueness of each validate the significance of the choice of our model and framework.

Finally, Dawanage et al. [6] utilized the BlazePose model developed by Google, sharing similarities with our research approach in terms of real-time detection and efficiency. However, their use of a two-step process involving a detector and an estimator, in contrast to the more streamlined process facilitated by MediaPipe in our study, might result in additional computational overhead.

In summary, the comparison to these studies demonstrates the strength of this research approach with the use of the MediaPipe framework, offering distinct advantages in terms of efficiency, accuracy, and scalability for human pose estimation tasks. Furthermore, the chosen classifier in this study has proven to be an excellent fit, offering robust performance that upholds against the surveyed methodologies. Hence, this evaluation assures that this research approach presents a compelling methodology for human pose estimation, underpinning the validity and significance of these research results.

V. Conclusion

In this study, exercise postures for workouts performed at home are correctly identified and corrected using machine learning software. While it does an excellent job of handling different body angles, there are still several drawbacks, including the need for more varied exercises, improved feedback accuracy, and reliable functioning in a variety of settings.

These findings have important public health implications because they offer a practical fitness option that is essential for reducing sedentary behavior and promoting regular physical exercise. Additionally, it has the potential to effectively prevent injuries brought on by incorrect posture, thus enhancing the safety of exercising at home.

Moving forward, there is room for further growth and study to realize the software's full potential. Potential improvements include expanding the supported exercises, expanding feedback accuracy, and improving performance under various conditions.

Conclusively, this study has the potential to be a solid stepping point to revolutionise home training routines and make a substantial contribution to the promotion of regular physical activity and healthier lifestyles.

Appendices

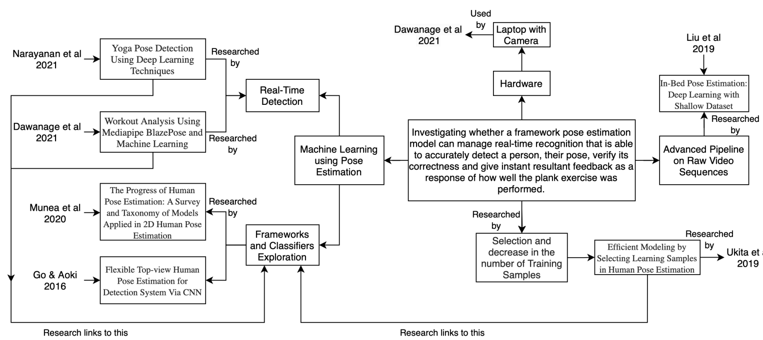

Fig. 4. Literature Map

References

1. Narayanan, S. S., Misra, D. K., Arora, K. & Rai, H. Yoga Pose Detection Using Deep Learning Techniques. *SSRN - Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3842656# (2021).
2. Go, R. & Aoki, Y. *Flexible top-view human pose estimation for detection system via CNN* in (2016), 1–4. https://ieeexplore.ieee.org/document/7800406.
3. Liu, S., Yin, Y. & Ostadabbas, S. In-Bed Pose Estimation: Deep Learning With Shallow Dataset. *IEEE Journal of Translational Engineering in Health and Medicine* **7,** 1–12. ISSN: 2168-2372. https://ieeexplore.ieee.org/document/8611350?denied= (2019).
4. Ukita, N., Matsuyama, Y. & Hagita, N. *Efficient modeling by selecting learning samples in human pose estimation* in (2014), 1–6. https://ieeexplore.ieee.org/document/6958917.
5. Munea, T. L. *et al.* The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation. *2022 IEEE World Conference on Applied Intelligence and Computing (AIC)* **8,** 133330–133348. ISSN: 2169-3536. https://ieeexplore.ieee.org/document/9144178 (2020).
6. Dawange, S., Chavan, A. & Dusane, A. Workout Analysis Using Mediapipe BlazePose and Machine Learning. *International Journal Of Creative Research Thoughts (IJCRT)* **9,** 2320–2882. https://ijcrt.org/papers/IJCRT2112453.pdf (2021).