Indonesian Journal of Electrical Engineering and Computer Science
Vol. 32, No. 1, April 2025, pp. 555∼568
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v32.i1.pp555-568

HorseNet: a novel deep learning approach for horse health classification

Nesrine Atitallah[1], Ahmed Abdel-Wahab[1], Anas A. Hadi[1], Hussein Abdel-Jaber[1], Ali Wagdy Mohamed[2], Mohamed Elsersy[3], Yusuf Mansour[1]
[1]Faculty of Computer Studies, Arab Open University, Riyadh, Saudi Arabia
[2]Department of Operations Research, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt
[3]Department of Computer Information Systems, Higher Colleges of Technology, Al Ain, Abu Dhabi, United Arab Emirates

Article Info

Article history:

Received Jan 28, 2024
Revised Oct 15, 2024
Accepted Oct 30, 2024

Keywords:

Convolutional neural networks
Deep learning
Horse wellness classification
Inception
Transfer learning
VGG16

ABSTRACT

In equestrian sports and veterinary medicine, horse welfare is paramount. Horse tiredness, lameness, colic, and anemia can be identified and classified using deep learning (DL) models. These technologies analyze horse images and videos to help vets and researchers find symptoms and trends that are hard to see. Early detection and better treatment of certain disorders can improve horses' health. DL models can also improve with new data, improving diagnosis accuracy and efficiency. This study comprehensively evaluates three convolutional neural network (CNN) models to distinguish normal and abnormal horses using the generated horse dataset. For this study, a unique dataset of horse breeds and their normal and abnormal states was collected. The dataset includes mobility patterns from this study's initial data collection. DL models like CNNs and transfer learning (TL) models (visual geometry group (VGG)16, InceptionV3) were employed for categorization. The InceptionV3 model outperformed CNN and VGG16 with over 97% accuracy. Its depth and multi-level structure allow the InceptionV3 model to recognize characteristics in images of varied scales and complexities, explaining its excellent performance.

Corresponding Author:

Ahmed Abdel-Wahab
Faculty of Computer Studies, Arab Open University
Riyadh, Saudi Arabia
Email: a.rakha@arabou.edu.sa

1. INTRODUCTION

The surveillance of animal health, specifically in horses, is of utmost importance in guaranteeing their welfare, productivity, and economic significance in both professional and leisure contexts [1]. Conventional approaches to monitor horse health often include labor-intensive, subjective, and time-consuming manual observation and assessment conducted by veterinarians or caregivers [2]. There are numerous inherent limitations associated with traditional techniques for monitoring the health of horses. Initially, it is important to note that the process of manually observing and evaluating the health of horses by veterinarians or caretakers is often characterized by a significant amount of labor and time expenditure [3]. However, shifts in the expertise and background of those caring for or treating the horse can affect the reliability of these evaluations. This can lead to inconsistencies when judging the animal's overall well-being. This may lead to inconsistencies when judging the animal's overall well-being. Furthermore, it is important to note that these methodologies may exhibit limitations in identifying nuanced alterations in individuals' health statuses, including the first indications of ailments or injuries. If left undetected and untreated in a timely manner, these illnesses have the potential to deteriorate progressively. These issues have the potential to undermine the efficacy of conventional procedures, hence emphasizing the need for new methodologies that may effectively tackle these limits and enhance the precision and efficiency of horse health monitoring [4]. The use of cutting-edge technology, such as artificial intelligence (AI), presents novel prospects for the automation and enhancement of animal health monitoring, hence enabling a more complete and non-intrusive approach to evaluate animal well-being. Through the adoption of novel approaches, it is possible to augment the well-being, efficacy, and economic worth of horses, guaranteeing their sustained prosperity in both leisure and occupational contexts [5].

These technologies have the capability to automate and improve the precision of animal health monitoring, therefore offering a more complete and non-intrusive approach to evaluating animal well-being. An illustration of the use of wearable sensors is their utilization in the continuous monitoring of essential physiological indicators, including heart rate, breathing rate, and temperature. These sensors possess the capability to communicate data in real-time to a system based on cloud computing [6]. This system facilitates the analysis of the data via the use of machine learning (ML) algorithms, thereby enabling the identification of the first indications of disease or damage. In a similar vein, computer vision algorithms may be used to examine video recordings of horses to identify alterations in their gait or behavior, which may serve as potential indicators of underlying health conditions [7]. These advanced technologies can improve horse health monitoring accuracy, efficiency, and efficacy. Technology can increase animal well being and benefit horse owners, caregivers, and vets. They can provide timely and accurate horse health information to help them make informed horse care and treatment decisions [8].

Deep learning (DL) has the potential to profoundly transform the field of animal health monitoring. With extensive datasets, DL algorithms have the capacity to acquire knowledge about patterns and then provide precise forecasts pertaining to the well-being of animals [9]. This technological advancement has several advantages in comparison to conventional approaches for monitoring animal well-being, including enhanced precision, automation, and timely identification of health concerns. The use of DL techniques enables the continuous and non-invasive monitoring of animal health, hence facilitating a more thorough evaluation of an animal's overall health condition. Furthermore, DL has the capability to automate the process of monitoring, therefore mitigating the labor-intensive characteristics associated with conventional monitoring techniques [10]. The timely identification of health concerns is of utmost importance in mitigating the progression of severe medical ailments. In this regard, the use of DL algorithms may play a pivotal role in promptly detecting small alterations in an animal's health, hence facilitating timely intervention and treatment. These advantages illustrate the capacity of DL to significantly transform the field of animal health monitoring and enhance the welfare of animals [11]. Transfer learning (TL) is a method of reusing a model that has already been trained on one task to solve a new, related task. This can save a lot of time and effort, and often leads to better results than training a new model from scratch, especially when there is limited data available. In fact, several works used TL to classify images. Noor *et al.* [12] proposed a dataset of sheep facial images and a framework that leverages TL and fine-tuning to automatically differentiate between images of sheep faces showing pain and those that appear normal. In addition, Tammina [13] employed the VGG-16 model, which is a pre-trained deep CNN, to classify images. This study presents an innovative methodology for the identification of horse health status via the use of DL methodologies. This study introduces several key innovations in horse health monitoring:

i) HorseSet 1.0: a comprehensive dataset of horse wellness images, capturing diverse situations, weather conditions, lighting, and angles.
ii) Expert-validated data: annotated and reviewed by veterinary specialists with over a decade of experience, ensuring reliability and accuracy.
iii) HorseNet: an efficient DL approach for rapid and accurate detection and classification of horse wellness.
iv) Performance evaluation: implementation and assessment of multiple DL algorithms (VGG16, InceptionV3, proposed CNN) using the novel dataset.

This work is structured in the following manner: section 2 provided a thorough examination of the current body of research pertaining to the use of DL algorithms in horse monitoring. Section 3 of the document provides a comprehensive exposition of the proposed model, including a thorough depiction of the dataset and the DL methodologies used within it. Section 4 presents the results of the study and how they were found. It includes any statistical analyses or pictures of the data, like the F1-score, accuracy, confusion matrix, loss, precision, recall, and loss. Section 5 encompasses an analysis and interpretation of the obtained results. This includes a comprehensive discussion on the implications of the findings for the proposed model as well as a thorough performance study of the model's outcomes. Additionally, this part provides a concise comparison of the DL technologies included in the proposed model. The last part presents a concise overview of the primary discoveries and ramifications of the investigation. Furthermore, this part provides an analysis of potential avenues for further study.

2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.1. Literature review

The application of ML and DL in animal health monitoring has gained momentum, with studies focusing on various species, including horses. For instance, Pallathadka *et al.* [14] demonstrated the effectiveness of DL in identifying lameness in horses through video analysis, achieving a 91% accuracy, outperforming traditional methods. DL has also been applied to detect colic in horses using data from wearable sensors, showing 90% precision [15], and has provided insights into horse behavior, fitness, and athletic abilities [16].

In pig health monitoring, Kavlak *et al.* [17] employed ML (extreme gradient boosting (XG-boost)) to predict health issues by analyzing feeding behaviors. Although the approach showed good sensitivity and specificity, it faced challenges due to unbalanced data and low symptom prevalence, emphasizing the need for better data quality and swine management strategies [17]. Similarly, TL, which adapts pre-trained models for new tasks, has been explored for wildlife monitoring by Nguyen *et al.* [18]. Their DL-based system showed promise in identifying animals in images, even with imbalanced datasets. The study suggests further enhancements through augmented datasets and advanced CNN models. Thermal image processing combined with ML has been used to detect avian diseases. Sadeghi *et al.* [19] achieved high accuracy using support vector machines (SVM) and artificial neural networks (ANN) to classify diseases like newcastle disease and avian influenza in broilers. After optimization, the SVM with Dempster–Shafer evidence theory outperformed ANN, with over 97% accuracy in disease classification [19].

Another study by Quaderi *et al.* [20] applied DL to detect beehive sounds on their own datasets and using various methods to reduce features. Sequential neural networks with AdaMax and sigmoid activation functions performed well, outperforming other methods. Random forests were also effective. When combining different types of data, sequential neural networks again showed the best results. Recurrent neural networks were particularly good at distinguishing bee sounds from noise [20]. DL has also been integrated with internet of things (IoT) technology for animal care and surveillance. Patil and Ansari [21] developed an intelligent system using CNNs and recurrent neural networks (RNNs) to monitor animal health, recognize activities, and detect environmental anomalies. The study highlights the potential of DL to improve animal welfare through proactive care. Sreedevi and Anitha [22] focused on using DL for wildlife detection via images and videos, showing that CNN-based approaches are effective for identifying different wild animal species, with practical implications for wildlife conservation. The welfare of animals has become a critical concern, leading to the development of wearable health monitoring devices using IoT. These devices collect vital signs such as body temperature, heart rate, and respiration rate, which can be transmitted to veterinary professionals for timely intervention. The cattle industry, in particular, could benefit from such systems, which allow continuous monitoring of individual animals' welfare. A study described a telemonitoring system prototype using wearable technology to enhance decision-making and improve horse welfare. Digital tools have the potential to enhance equine health monitoring by improving the accuracy and efficiency of health assessments [23]. While there have been significant advances in using DL for equine health monitoring, there are still areas that need improvement. Research has primarily focused on detecting lameness and colic in horses, with limited studies exploring respiratory and metabolic issues. DL requires large, high-quality datasets, making data acquisition and standardization challenging. Additionally, user-friendly tools are needed to integrate DL into health monitoring practices. Despite these challenges, DL offers promising opportunities to enhance horse welfare, performance, and economic value. Future research should address these gaps and develop practical tools for equine health monitoring [24]. To gain a comprehensive understanding of the existing approaches for horse health classification, we conducted a detailed overview of the various methodologies employed in prior studies as illustrated in Table 1. This review provides an overview of various methodologies used in horse health classification, highlighting dataset characteristics, classification algorithms, and performance metrics. The analysis serves as a valuable resource for researchers and identifies gaps and opportunities for further improvement in horse health monitoring.

Table 1. Comparison of related works

Reference No.	Year	Domain	Algorithm Used	Problem to be solved	Performance
[14]	2023	Identification of horse lameness	DL	Lameness identification of horse	Accuracy of 91%
[15]	2023	Analysis from wearable sensors to detect colic issue in horse	DL	Detect of the COVID-19 virus from wearable sensors	Precision 90%
[16]	2023	The health and fitness of horses	CNN	Monitoring the horses using gait analysis	DeepLabCut DLC Ver2.2 tool used
[17]	2023	Enhanced data quality	XGBoost	Data quality enhancement	Animal diseases management
[18]	2017	Automated wildlife monitoring	CNN TL	Detection and identification of the classified species species	Wildlife spotter image dataset, accuracy of 96.6%
[19]	2023	Classification of newcastle disease	SVM with Dempster–Shafer evidence theory	Efficacy of thermography and ML techniques in the classification of newcastle disease and avian influenza among broilers	Accuracy of over 97%
[20]	2022	Beehive sound analysis	CNN, RNN	Classify bee sounds from the non-beehive noises	AdaMax optimizer, accuracy of 85%
[21]	2020	Smart surveillance of stray animals	CNN, RNN	Monitor stray dog dog animals in a particular area	Accuracy of 85-90 %
[22]	2022	Wild animal classification	CNN, rectified linear unit (ReLU)	Wildlife animal detection IWildCam dataset	Accuracy of 87.8%
[24]	2023	Integration of DL monitoring for horse health surveillance	YOLO v7	A hybrid technique of automated muzzle feature extraction empowered by Yolo and SIFT, and feature matching using FLANN	Accuracy of 96.2%

2.2. Transfer learning convolutional neural networks

TL involves leveraging a model pre-trained on one problem to solve a different problem. This approach is gaining traction in the realm of deep neural networks, given their substantial data and computational demands. It allows for a more efficient training process by refining a previously trained DL model for a similar task. Essentially, it harnesses the insights a model has garnered from a data-rich task and applies it to a new task with limited data. To circumvent the challenges of training duration and data volume, we employed TL on two distinct pre-trained CNN models: VGG16, pioneered by Simonyan and Zisserman [25], and InceptionV3, crafted by Google's research team.

2.2.1. VGG16

The VGG16 model, developed by the visual geometry group at Oxford University, is a renowned neural network architecture for image classification. Its simple yet efficient design comprises 16 layers (13 convolutional and 3 fully connected), enabling the extraction of hierarchical features from images. The model's depth allows it to learn complex image representations, making it versatile for various visual recognition tasks. VGG16's straightforward structure and high performance have made it influential in advancing DL and computer vision, cementing its popularity among researchers and practitioners in the field [26].

2.2.2. InceptionV3

InceptionV3, developed by Google, is a sophisticated CNN architecture designed for large-scale image recognition and classification. Its innovative design features inception modules that capture information at multiple scales, striking an optimal balance between depth and computational efficiency. This approach enables high accuracy in image classification tasks while maintaining scalability for large datasets. InceptionV3's versatility has led to its widespread adoption across various domains, consistently achieving top-tier results in benchmark competitions. Its combination of performance and efficiency makes it a popular choice for complex image recognition tasks, particularly when processing extensive datasets [27]. Due to its impressive performance and efficiency, the InceptionV3 model has become a reference architecture in the field of DL and image classification. It has served as a foundation for subsequent advancements in CNN architecture and has influenced the development of newer models. Researchers and practitioners alike continue to leverage the capabilities of InceptionV3 to tackle complex image recognition challenges and push the boundaries of computer vision applications.

3. METHODS AND MATERIALS

3.1. Dataset

In order to effectively classify and detect the well-being of horses using camera and video footage, we utilized advanced computer vision techniques and DL algorithms. This innovative approach allowed us to develop a system capable of analyzing visual data and extracting meaningful insights related to the health and condition of horses. In the upcoming section, we will introduce and elaborate on our meticulously curated dataset, known as HorseNet. This dataset plays a crucial role in our research as it serves as the foundation for training and evaluating the performance of our models. The subsequent discussion will encompass a detailed description of the dataset, including its composition, size, and the specific attributes and features captured within. By shedding light on the intricacies of HorseNet, we aim to provide a comprehensive understanding of the underlying data that powers our horse wellness classification and detection system.

3.1.1. Study area

The research study was conducted at the esteemed horse club located in Ahsa, a region known for its rich equestrian culture. During the initial stages of the research, the authors encountered a challenge in finding a suitable dataset that adequately represented the diverse range of horses and conditions relevant to their study. To overcome this hurdle, the authors took the initiative to gather and curate the necessary data themselves. They meticulously collected a comprehensive set of images featuring horses, ensuring that it encompassed a wide spectrum of breeds, ages, and physical attributes. These images were specifically sourced from the horse club situated in the captivating Ahsa region of Saudi Arabia, which boasts a notable reputation for its dedication to the equestrian arts. By actively engaging with the horse club and obtaining their cooperation, the authors were able to acquire a diverse and representative dataset that formed the foundation of their research.

3.1.2. Data collection

Our research endeavors led us to amass a vast and comprehensive collection of horse images and videos, encompassing a diverse range of breeds, ages, and varying states of health conditions. To accomplish this, we established a fruitful partnership with knowledgeable members of the esteemed horse club. Working in collaboration, we utilized smartphones equipped with high-quality cameras to capture the visual data. The convenience and portability of these devices allowed us to efficiently document the horses in their natural environments, ensuring the authenticity and representatives of the collected media. Figure 1 illustrates the process, showcasing the use of smartphones in capturing the images and videos that constitute our extensive dataset. Through this diligent and collaborative effort, we were able to create a rich and diverse resource that forms the backbone of our research on horse wellness.

Figure 1. Horse wellness dataset samples

The pictures were stored as PNG files in multiple resolutions. Notably, the primary focus of the dataset is on images depicting horses in diverse settings. These photos were gathered from January to April 2023 and were influenced by different weather and illumination conditions. They were also shot from a range of perspectives. Out of the entire collection, only 1,218 pictures met the criteria for the research. The rest, which were either out of focus or didn't portray the necessary scenarios involving horses, were excluded.

3.1.3. Horse wellness classifications

DL techniques, specifically CNNs, have shown great potential in various image recognition tasks. While some aspects of fatigue, lameness, colic, and anemia in horses can potentially be detected using images and DL. In the following, there are some aspects that might be detected using images and DL:

i) Posture and gait analysis: DL algorithms can analyze images or videos of a horse's posture and gait to detect abnormalities, such as limping, stiffness, or uneven weight distribution. This analysis can help classify the severity and type of lameness.
ii) Physical appearance: fatigue, anemia, and colic can cause changes in a horse's physical appearance that may be detectable using DL. For example, an anemic horse may exhibit pale mucous membranes (gums), and a horse with colic may show signs of discomfort or abdominal distension. DL models can be trained to recognize these features from images and identify horses that may need further examination.
iii) Behavioral patterns: horses suffering from fatigue, lameness, colic, or anemia may exhibit abnormal behaviors such as restlessness, rolling, or frequent changes in position. By analyzing a series of images or videos, DL models may be able to detect these behavior patterns and help in identifying affected horses.

Table 2 presents a comprehensive categorization of horse wellness, outlining six distinct classes of horse behavior and health. It ranges from the "Normal Horse" category, which comprises the majority with 178 horses, to more specific behaviors like "Rolling Horse" and "Stretching Horse".

Table 2. Horse wellness categories

Wellness class	Class name	Number
Normal horse	0	478
Rolling horse	1	259
Stretching horse	2	21
Pawing	3	358
Lying down	4	38
Biting at sides	5	62
Kicking belly	6	2

3.1.4. Data distribution

In Table 3, we present a comprehensive overview of the horse wellness datasets used in our study. Upon analyzing these datasets, we observed an inherent imbalance in the distribution of instances across the various horse wellness categories. This imbalance posed a challenge as it could potentially bias the performance of our DL algorithms during training. To address this issue and promote fairness in our model training, we employed image enhancement methods. These techniques allowed us to manipulate and augment the dataset, ensuring a more balanced distribution of instances across the different horse wellness categories. By equalizing the representation of each category, we aimed to enhance the algorithm's ability to learn and generalize patterns associated with the full spectrum of horse wellness. Through the application of these image enhancement methods, we aimed to mitigate any potential biases that could arise due to the imbalanced nature of the original dataset, ultimately fostering more accurate and reliable results.

Table 3. Distribution in our dataset

No.	Class name	Number of samples
1	Normal horse	478
2	Abnormal horse	740
	Total	1,218

3.2. Methods

The proposed detection and classification framework entails a well-defined workflow comprising five key steps: gathering data, refining, and augmenting this data, employing various CNN models, and conducting experiments and assessments as depicted in Figure 2. These steps collectively contribute to the accuracy and effectiveness of the framework.

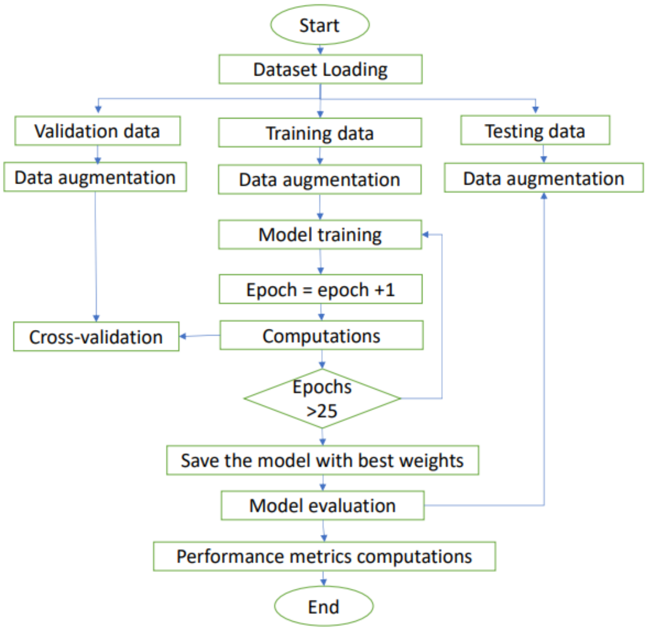


Figure 2. The experimentation workflow

3.2.1. Data pre-processing

A critical phase in any computer vision (CV) system involves prepossessing the images. Initially, every image was resized to dimensions of 300×300 to ensure square-shaped images and consistency throughout the dataset. Subsequently, these images were altered to match the input dimensions of various models. For the custom designed HorseNet, the images were adjusted to 256×256 pixels. For VGG16 and traditional CNN models, they were changed to 224×224 pixels. Meanwhile, the input for InceptionV3 was resized to 71×71 pixels.

3.2.2. Data augmentation

DL models typically need a vast amount of data for optimal performance. When there's limited training data available, image augmentation is often used to bolster the robustness of image classifiers. Image augmentation artificially produces training samples using techniques like rotation, noise addition, shifting, mirroring, and blurring. From the initial dataset of 1,218 images, after augmentation, it was split: 90% was used for training and validation, and 10% for testing, as illustrated in Table 4. In this study, we employed four augmentation methods: rotation, flipping, zoom, and brightness adjustments. Rotating images is a common technique, expanding the dataset by producing variants of the original images rotating anywhere from 0 to 360 degrees. Flipping, on the other hand, can be seen as a subset of rotation, creating mirrored versions of the original.

Table 4. Distribution of images dataset after data augmentation

	Normal	Abnormal	Total
Training data	6,920	9,740	16,658
Validation data	2,488	2,170	4,658
Testing data	1,485	1,476	2,961
Total	10,893	13,386	24,277

3.3. Model design

3.3.1. Proposed CNN

This subsection tackled the architecture of the proposed traditional CNN designed for the classification of horse wellness. The model is composed of six main components: 10 convolutional layers paired with 4 pooling layers, two fully connected layers, and a ReLU activation layer. Different filters were utilized in each convolution layer to extract varied features [28]. To counteract over fitting, dropout was integrated as a regularization technique within both the max-pooling and the fully connected layers. The input images had dimensions of 32×32×3. By choosing a batch size of 64 and a learning rate of 0.0001, the training speed was optimized. Two deep layers were established: the inaugural layer took an input channel of one, featuring a 3×3 kernel, a stride of one, and a padding value of two. Post-convolution, the image dimensions shrank; however, padding was adjusted to zero to retain the original size. The ReLU function was favored as the model's activation due to its resilience against saturation and its gradient performance relative to other activation. The max-pooling layer employed both a kernel and stride of two. For the subsequent convolutional layer, the input dimension was 32, output was 64, using a 5x5 kernel, a stride of one, and padding of two. The ReLU and max pooling remained consistent across both layers. A dropout layer was integrated to alleviate and mitigate over fitting. Conclusively, two fully connected layers ensured the interlinking of all neurons. The CNN's blueprint can be visualized in Figure 3.

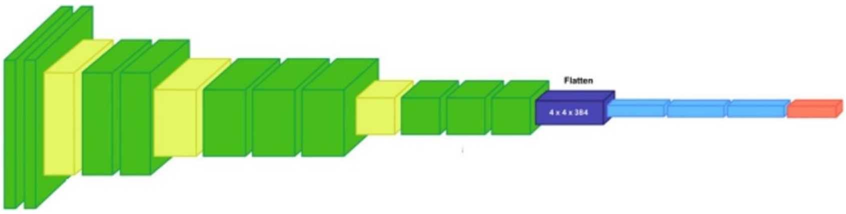


Figure 3. Architecture of the proposed CNN

3.3.2. Proposed approach

In this research, we utilized three distinct models: a custom-built CNN, VGG-16, and InceptionV3. We trained our models using augmented data from the primary dataset, as detailed in section 3.1.. Post-augmentation, the entire dataset, consisting of 24,277 images, was split into 16,658 training, 4,658 validation, and 2,961 test images, as illustrated in Table 4. We adopted TL method to address the challenges of extended training times and limited data availability. Moreover, the top layers, along with the gully connected (FC) layers added to the tail-end of the pre-existing models, were set to a static state, and subsequently retrained on our specific dataset to achieve the targeted outcomes. In our methodology, we thawed the concluding layers of these established models, then retrained them using our dataset, while keeping the preliminary layers frozen. This process is visualized in Figure 4.

4. RESULTS AND DISCUSSION

In this section, we will provide a comprehensive summary of the results obtained from the experiments conducted on the dataset pertaining to horse well-being. These experiments were conducted with the aim of utilizing advanced CV techniques and DL algorithms to gain insights into the health and condition of horses. The collected dataset, comprising a diverse collection of horse images and videos, formed the basis for the experiments. Using this dataset, we trained and evaluated several CNN models specifically designed for horse well-being analysis. These models were selected based on their established performance in image classification tasks and their suitability for the domain of horse well-being. Through rigorous experimentation and analysis, we obtained valuable findings regarding the classification and detection of horse well-being using Dl techniques. The results shed light on the efficacy of the employed CNN models in accurately assessing various aspects of horse well-being, such as overall health, body condition, and any potential signs of discomfort or distress. Furthermore, in order to provide a comprehensive analysis, we compared the performance of the different CNN models used in the experiments. This comparative analysis allowed us to identify the strengths and weaknesses of each model, enabling us to make informed decisions regarding their suitability for specific horse well-being assessment tasks. By presenting these summarized results and conducting a thorough analysis, we aim to contribute to the existing body of knowledge utilizing CV and DL for horse well-being analysis. The findings obtained from these experiments have the potential to advance the understanding of horse health assessment, potentially leading to improved care and well-being for these magnificent animals.

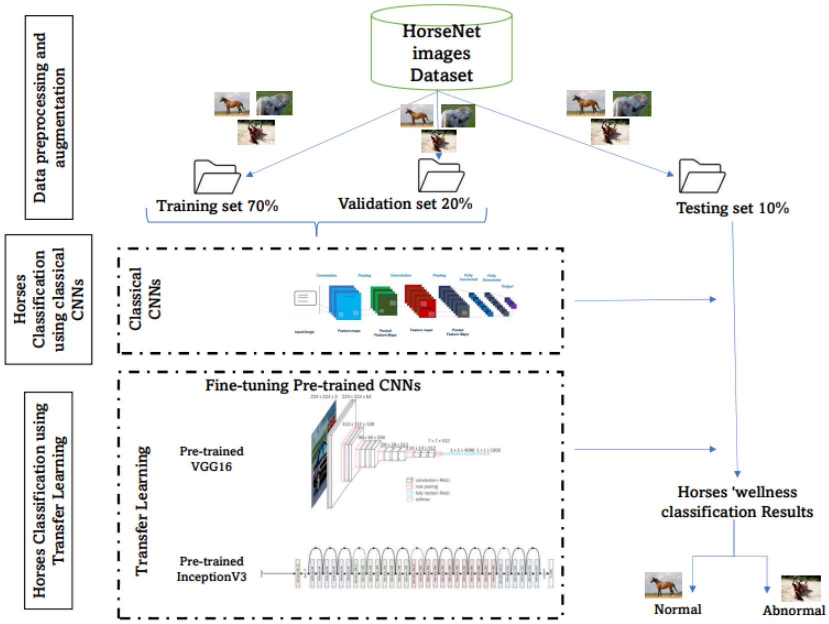


Figure 4. Proposed approach for horse wellness detection and binary-classification

4.1. Experimental setup

In our research, we employed the Jupiter notebook to script the entire workflow using Python 3.8. To build the neural networks, we incorporated both the Keras library and TensorFlow as the backend. Furthermore, OpenCV facilitated data loading and pre-processing, while Sci-Kit Learn was instrumental in generating classification summaries. For expedited computational performance, we utilized the Nvidia GeForce MX 250 GPU, supplemented by CUDA and cuDNN libraries. cuDNN is a GPU-enhanced library tailored to boost various DL frameworks. The system underpinning our work had these specifications: a 64-bit processor, an Intel Core i7-8565U CPU clocking at 1.80 GHz, and 16 GB of RAM. It operated on a Windows 10 platform equipped with NVIDIA GeForce MX.

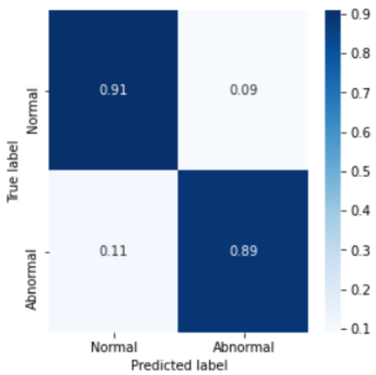
4.2. Experimental results

The main objective of the suggested approach is to accurately identify and categorize the health of horses. To achieve this, we utilized three distinct CNN models. Table 5 presents the performance metrics of the CNN models post fine-tuning and training on the horse wellness dataset. Among the models, InceptionV3 stands out with commendable accuracy, precision, and recall rate, each at 97%. This is trailed by VGG16 and subsequently the traditional CNN. Overall, the models grounded in TL showcased robust performance, boasting an accuracy surpassing 95%. InceptionV3 emerges as the top performer with all metrics at or near 97%. This indicates not only high accuracy but also a balanced capability in precision and recall, suggesting a few false positives and false negatives. This balance is crucial in medical or wellness contexts where both types of errors carry significant consequences. VGG16 follows closely, demonstrating that while slightly less effective than InceptionV3, it still provides a highly reliable method for classifying horse health, with metrics around 95%. Proposed CNN, while trailing behind the other two models, still shows respectable performance metrics (90%). This indicates a viable option when computational resources are limited or for preliminary explorations.

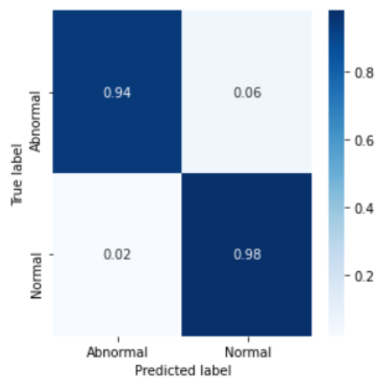
Table 5. Performance metrics of the deployed TL-based models and the proposed CNN

Model	Accuracy	Precision	Recall	F1 score
Proposed CNN	90.20%	90%	90%	90%
VGG16	96.18%	96%	96%	96%
InceptionV3	96.92%	97%	97%	97%

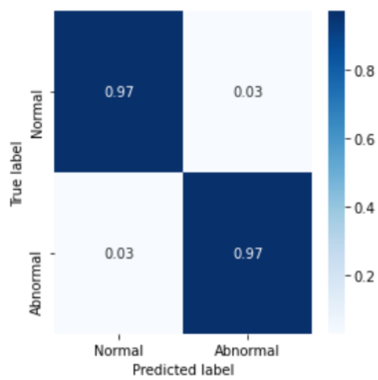
Figure 5 shows the plots of confusion matrices for each horse wellness class produced by the proposed CNN, VGG16, and InceptionV3 models. Figure 5(a) represents the model with the lowest performance, demonstrating greater difficulty in distinguishing between normal and abnormal conditions compared to the others. Figure 5(b) shows marked improvement in both sensitivity and specificity, with a balanced decrease in both false positives and false negatives, suggesting better overall performance. Figure 5(c) stands out as the best performer, with the highest true positive and true negative rates, coupled with the lowest false positive and false negative rates, indicating a highly accurate model. The progression observed is consistent with the general expectation that more advanced models with more parameters and sophisticated architectures, often utilizing TL, will generally outperform simpler models, particularly on complex tasks like image classification.



(a)



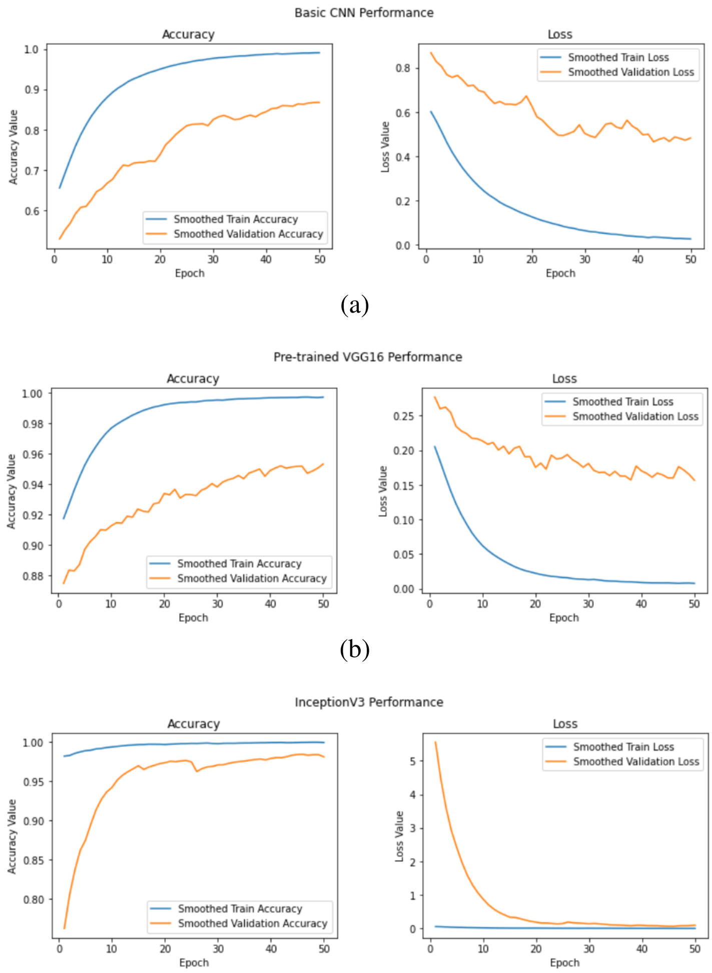
(b)



(c)

Figure 5. The confusion matrix of: (a) the proposed CNN model, (b) the pre-trained VGG16, and (c) the pre-trained InceptionV3 approach

The analysis of curves depicted in Figures 6, which includes the smoothed training curves and validation loss and accuracy curves of the proposed CNN model in Figure 6(a), the ore-trained VGG16 approach in Figure 6(b), and the pre-trained InceptionV3 approach in Figure 6(c). This Figure suggests that the pre-trained InceptionV3 model is the best performer in terms of learning effectively and generalizing from the training data to the validation data. It also manages to achieve high accuracy while avoiding over-fitting, making it the most suitable model for deployment in real-world scenarios where model robustness and reliability are crucial.



(a) (b) (c)

Figure 6. The smoothed train and validation loss and accuracy curves of: (a) the proposed CNN model, (b) the pre-trained VGG16 approach, and (c) the pre-trained InceptionV3 approach

4.3. Discussions

This study aims to investigate the use of TL to detect and classify horse wellness. Three TL techniques have been used, namely InceptionV3 and VGG16. Comparing with proposed CNN results reported with the same dataset, as shown in Table 5, we can conclude that TL outperforms proposed CNNs where the accuracy is 90%. The progression from the proposed CNN to the more sophisticated pre-trained models demonstrates the benefit of using advanced architectures and pre-training on large datasets. In fact, pre-trained models are easy to work with and train due to retained knowledge from previous learning and they can achieve good performance compared with proposed CNNs. Furthermore, TL offers faster training as pre-trained models are trained on large datasets, often with powerful hardware and extensive time. Therefore, we can leverage the knowledge learned by these models and avoid training from scratch. This can significantly reduce training time and computational resources required. Moreover, TL requires less data. Indeed, DL models typically require a large amount of data to generalize well. The integration of our novel dataset and smart detection algorithms represents a significant advancement in equine health research. This innovative approach combines cutting-edge DL techniques with specialized equine health data, addressing a critical gap in the field. Our dataset, curated to capture diverse aspects of horse wellness, provides a comprehensive foundation for developing more accurate and nuanced detection models. The smart detection algorithms we have implemented, particularly through TL, demonstrate remarkable efficacy in identifying subtle health indicators that may be overlooked by traditional methods. This breakthrough has the potential to revolutionize preventive care in equine medicine, enabling earlier interventions and more personalized treatment strategies. Furthermore, our research contributes to the broader field of veterinary informatics by showcasing the potential of advanced AI techniques in animal health monitoring. The methodologies developed in this study could potentially be adapted for other large animal species, opening new avenues for research in comparative medicine.

Our dataset can be expanded to encompass a wider range of horse breeds, ages, and health conditions across diverse geographical locations. Moreover, clinical validation studies should be conducted in multiple equine veterinary settings to compare our AI system's performance with diagnoses from experienced veterinarians. Further algorithm refinement, economic impact assessments, and user acceptance studies would contribute to improving the system's accuracy, cost-effectiveness, and practical implementation. These research directions aim to validate, expand, and refine our initial findings, potentially leading to significant advancements in AI-assisted equine healthcare.

5. CONCLUSIONS

This study examined the use of DL algorithms to monitor horse health and identify normal and abnormal animals. We used a well-chosen dataset from horse clubs with a variety of breeds and health concerns. Pre-trained InceptionV3 surpassed CNN and other pre-trained models with an accuracy of over 97%. The study showed the benefits of TL with pre-trained models. Working with and training these models is easier since they remember past learning activities. For horse health monitoring, pre-trained models performed as well as traditional CNNs. Rapid training was another benefit of TL. Pre-trained models are trained on large datasets using sophisticated hardware and time, so we can use their knowledge instead of starting from scratch. This decreases training time and computational resources, making it realistic for real-world applications. TL also performed well with less data. DL models need lots of data to generalize. We can use pre-trained models to learn from large datasets and perform well with less data. This research advances horse health monitoring by demonstrating the efficacy of DL systems, particularly the InceptionV3 algorithm. The results show that these algorithms can help horses in equestrian activities by detecting and treating various health concerns early. Future research could expand upon this work by incorporating larger datasets and refining the DL techniques for enhanced treatment and management of equine health. Additionally, applying these models in practical scenarios holds considerable promise. This would entail integrating the algorithms into veterinary diagnostic frameworks and evaluating their efficacy based on tangible clinical results.

ACKNOWLEDGEMENTS

The authors extend their appreciation to the Arab Open University for funding this work through AOU research fund no. (AOURG-2023-012).

REFERENCES

[1] J. Zinsstag *et al.*, "Advancing one human–animal–environment health for global health security: what does the evidence say?," *The Lancet*, vol. 401, no. 10376, pp. 591–604, 2023, doi: 10.1016/S0140-6736(22)01595-1.

[2] A. B. Giles, P. Scanes, A. Dickson, B. Adam, and B. Kelaher, "Drones are an effective tool to assess the impact of feral horses in an alpine riparian environment," *Austral Ecology*, vol. 48, no. 2, pp. 359–373, 2023, doi: 10.1111/aec.13271.

[3] P. Bhosale and S. Jadhav, "An insight on methodologies used for predictive machine health monitoring: a review," *Application of Communication Computational Intelligence and Learning*, pp. 23–34, 2022, doi: 10.1201/9781003340867-3.

[4] M. Miller, R. Byfield, M. Crosby, and J. Lin, "Networked wearable sensors for monitoring health and activities of an equine herd: an IoT approach to improve horse welfare," *IEEE Sensors Journal*, 2024, doi: 10.1109/JSEN.2024.3436665.

[5] R. Ezhilarasie, N. Aishwarya, V. Subramani, and A. Umamakeswari, "Acceleration of computer vision and deep learning: surveillance systems," in *Handbook of Research on Computer Vision and Image Processing in the Deep Learning Era, IGI Global*, 2022, pp. 19–28.

[6] R. C. Qian and Y. T. Long, "Wearable chemosensors: a review of recent progress," *ChemistryOpen*, vol. 7, no. 2, pp. 118–130, 2018, doi: 10.1002/open.201700159.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.

[8] K. Džermeikaitė, D. Bačėninaitė, and R. Antanaitis, "Innovations in cattle farming: application of innovative technologies and sensors in the diagnosis of diseases," *Animals*, vol. 13, no. 5, p. 780, 2023, doi: 10.3390/ani13050780.

[9] A. Holzinger, K. Keiblinger, P. Holub, K. Zatloukal, and H. Müller, "AI for life: trends in artificial intelligence for biotechnology," *New Biotechnology*, vol. 74, pp. 16–24, 2023, doi: 10.1016/j.nbt.2023.02.001.

[10] S. A. A. Shah, A. H. Emara, A. A. Wahab, N. A. Algeelani, and N. A. Al-Sammarraie, "Street-crimes modelled arms recognition technique employing deep learning and quantum deep learning," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 30, no. 1, pp. 528–544, 2023, doi: 10.11591/ijeecs.v30.i1.pp528-544.

[11] R. Arablouei, L. Wang, L. Currie, J. Yates, F. A. P. Alvarenga, and G. J. Bishop-Hurley, "Animal behavior classification via deep learning on embedded systems," *Computers and Electronics in Agriculture*, vol. 207, p. 107707, 2023, doi: 10.1016/j.compag.2023.107707.
[12] A. Noor, Y. Zhao, A. Koubaa, L. Wu, R. Khan, and F. Y. O. Abdalla, "Automated sheep facial expression classification using deep transfer learning," *Computers and Electronics in Agriculture*, vol. 175, p. 105528, 2020, doi: 10.1016/j.compag.2020.105528.
[13] S. Tammina, "Transfer learning using VGG-16 with deep convolutional neural network for classifying images," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 9, no. 10, p. p9420, Oct. 2019, doi: 10.29322/IJSRP.9.10.2019.p9420.
[14] H. Pallathadka, M. Mustafa, D. T. Sanchez, G. Sekhar Sajja, S. Gour, and M. Naved, "Impact of machine learning on management, healthcare and agriculture," *Materials Today: Proceedings*, vol. 80, pp. 2803–2806, 2023, doi: 10.1016/j.matpr.2021.07.042.
[15] S. Broomé *et al.*, "Going deeper than tracking: a survey of computer-vision based recognition of animal pain and emotions," *International Journal of Computer Vision*, vol. 131, no. 2, pp. 572–590, Feb. 2023, doi: 10.1007/s11263-022-01716-3.
[16] N. Niknejad, J. L. Caro, R. Bidese-Puhl, Y. Bao, and E. A. Staiger, "Equine kinematic gait analysis using stereo videography and deep learning: stride length and stance duration estimation," *American Society of Agricultural and Biological Engineers*, vol. 66, no. 4, pp. 865–877, 2023, doi: 10.13031/ja.15386.
[17] A. T. Kavlak, M. Pastell, and P. Uimari, "Disease detection in pigs based on feeding behaviour traits using machine learning," *Biosystems Engineering*, vol. 226, pp. 132–143, 2023, doi: 10.1016/j.biosystemseng.2023.01.004.
[18] H. Nguyen *et al.*, "Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring," in *Proceedings - 2017 International Conference on Data Science and Advanced Analytics, DSAA*, 2017, vol. 2018-January, pp. 40–49, doi: 10.1109/DSAA.2017.31.
[19] M. Sadeghi, A. Banakar, S. Minaei, M. Orooji, A. Shoushtari, and G. Li, "Early detection of avian diseases based on thermography and artificial intelligence," *Animals*, vol. 13, no. 14, p. 2348, 2023, doi: 10.3390/ani13142348.
[20] S. J. S. Quaderi, S. A. Labonno, S. Mostafa, and S. Akhter, "Identify the beehive sound using deep learning," *International Journal of Computer Science and Information Technology*, vol. 14, no. 4, pp. 13–29, 2022, doi: 10.5121/ijcsit.2022.14402.
[21] H. Patil and N. Ansari, "Smart Surveillance and Animal Care system using IOT and Deep Learning," in *Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST)*, 2020, doi: 10.2139/ssrn.3565274.
[22] K. L. Sreedevi and A. Edison, "Wild animal detection using deep learning," in INDICON 2022 - 2022 IEEE 19th India Council International Conference, 2022, pp. 1–5, doi: 10.1109/INDICON56171.2022.10039799.
[23] P. Khatate, A. Savkar, and C. Y. Patil, "Wearable smart health monitoring system for animals," in *Proceedings of the 2nd International Conference on Trends in Electronics and Informatics*, ICOEI 2018, 2018, pp. 162–164, doi: 10.1109/ICOEI.2018.8553844.
[24] M. Ahmad, S. Abbas, A. Fatima, G. F. Issa, T. M. Ghazal, and M. A. Khan, "Deep transfer learning-based animal face identification model empowered with vision-based hybrid approach," *Applied Sciences (Switzerland)*, vol. 13, no. 2, p. 1178, 2023, doi: 10.3390/app13021178.
[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014, doi: 10.48550/arXiv.1409.1556.
[26] P. Gaur, H. Gupta, A. Chowdhury, K. McCreadie, R. B. Pachori, and H. Wang, "A sliding window common spatial pattern for enhancing motor imagery classification in EEG-BCI," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021, doi: 10.1109/TIM.2021.3051996.
[27] A. Phukan and D. Gupta, "EEG based emotion classification using xception architecture," in M*obile Radio Communications and 5G Networks: Proceedings of Second MRCN 2021*, vol. 339, Springer, 2022, pp. 95–108.
[28] O. Dahmane, M. Khelifi, M. Beladgham, and I. Kadri, "Pneumonia detection based on transfer learning and a combination of VGG19 and a CNN built from scratch," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 24, no. 3, pp. 1469–1480, Dec. 2021, doi: 10.11591/ijeecs.v24.i3.pp1469-1480.

BIOGRAPHIES OF AUTHORS

Nesrine Atitallah is an assistant professor at the Faculty of Computer Studies, Arab Open University, Saudi Arabia. She received the Polytechnic Engineering Diploma (with honors) in 2009, the M.Sc. degree in Electronic Systems and Communication Networks (with honors) in 2010 from the Tunisia Polytechnic School, and the Ph.D. degree in 2019 at the Computer and Embedded System Laboratory CES-Lab, Engineering National School of Sfax (ENIS), University of Sfax, Tunisia. Her research focuses on the development of energy aware reconfigurable node architecture for wireless sensor network. She is working on using intelligent models in cyber security, healthcare, and smart cities. She can be contacted at email: n.atitallah@arabou.edu.sa.

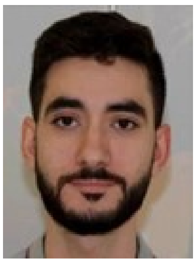
Ahmed Abdel-Wahab assistant professor, Faculty of Computer Studies, Arab Open University, Riyadh, Kingdom of Saudi Arabia. He holds a Ph.D. in Computer Engineering from the Department of Systems and Computer Engineering at Al-Azhar University in Cairo, Egypt since 2014. His research areas are data analysis and description, ML and DL algorithms, and e-learning. He can be contacted at email: a.rakha@arabou.edu.sa.

Anas A. Hadi is an assistant professor at Arab Open University (AOU), Riyadh. He received his Ph.D. degree in Computer Science in 2019 from Faculty of Computing and Information Technology, King Abdulaziz University (KAU). His current research interests include ML, optimization, brain computer interface BCI, and fature selection. He can be contacted at email: a.hadi@arabou.edu.sa.

Hussein Abdel-Jaber received his Ph.D. in Computing from University of Bradford in 2009, and from the same university, he received his M.Sc. in Mobile Computing in 2004. His research interests are in congestion control of networks (i.e. internet), queueing networks analysis using discrete-time queues or continuous-time queues, networks performance modelling and evaluation, fuzzy logic control, ML, data mining, computer security, and e-learning. He has several published research papers in the previously mentioned research interests. He can be contacted at email: habdeljaber@arabou.edu.sa.

Ali Wagdy Mohamed is a professor with Operations Research Department, Institute of Statistical Studies and Research, Cairo University. He is a prominent researcher with expertise in optimization, statistics, and data science. He has extensive experience in academia and industry, having held positions at prestigious universities and companies. He is a prolific author with over 180 publications and a highly sought-after reviewer for top scientific journals. His research interests span mathematical modeling, optimization techniques, and real-world applications in various fields. He can be contacted at email: aliwagdy@gmail.com.

Mohamed Elsersy Ph.D., lecturer, Computer, and Information Systems (CIS) Department, Al Ain Women's Campus, Higher Colleges of Technology, UAE; Mohamed Said Elsersy. (Member, IEEE) a Bachelor of Science in Computers and Systems Engineering (with first class honours) was awarded to him by Al-Azhar University in Cairo, Egypt. He received an M.Sc. in wireless communications engineering from the University of Calgary in Canada. The Memorial University of Newfoundland in St. John's, Newfoundland, Canada, awarded him his doctorate. He was a research fellow at Qatar University. The scholarship of Queen Elizabeth II has been bestowed upon him. Furthermore, he was honored with the best paper award at the 2nd International Conference on Emerging Technologies and Intelligent Systems (ICETIS 2022) for his contribution entitled "Privacy-preserving data aggregation scheme for e-health". Recent research centers on the application of ML algorithms to patient data and medical history to detect prevalent diseases like dementia at an early stage. Present research interests include the applications of ML to the transportation and healthcare industries. Subsequent research endeavors will center on the application of security and privacy consciousness to intelligent transportation systems and healthcare. He can be contacted at email: melsersy@hct.ac.ae.



Yusuf Mansour earned his bachelor's degree in information technology and computing from the Arab Open University in 2023. He is passionate about computer science, programming, and AI, particularly in ML and DL. His graduation project, "Horse health monitoring using artificial intelligence," focused on using camera-based horse movement analysis and ML to enhance horse health monitoring. This project showcased his technical skills and innovative problem-solving approach, highlighting his dedication to using technology for industry improvements. He can be contacted at email: yusef.sa98@gmail.com.