M. Shamim Hossain
Utku Kose
Deepak Gupta *Editors*

Explainable Machine Learning for Multimedia Based Healthcare Applications

Springer

Explainable Machine Learning for Multimedia Based Healthcare Applications

M. Shamim Hossain • Utku Kose • Deepak Gupta
Editors

Explainable Machine Learning for Multimedia Based Healthcare Applications

Springer

Editors
M. Shamim Hossain
Department of Software Engineering
College of Computer and Information Sciences
King Saud University
Riyadh, Saudi Arabia

Utku Kose
Department of Computer Engineering
Suleyman Demirel University
Isparta, Turkey

University of North Dakota
Grand Forks, ND, USA

Deepak Gupta
Maharaja Agrasen Institute of Technology
Delhi, India

ISBN 978-3-031-38035-8 ISBN 978-3-031-38036-5 (eBook)
https://doi.org/10.1007/978-3-031-38036-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

As the humanity, we have been facing many technological advancements with the start of the twenty-first century. Of course, these are somehow results of the previous discoveries and developments along the history, but the civilization feels the exponential growth and effects in especially last decade. Among all the technological advancements, the field of Artificial Intelligence has a critical role in developing not only present but also future world. The effects of intelligent systems have been even transforming the society. Today, one may say that all fields of the modern life is under dominance of intelligent systems as the supportive touches by the field of Artificial Intelligence. At this point, the healthcare has a critical place for sure. It is clear that Artificial Intelligence ensures revolutionary outcomes in the context of healthcare applications. Thanks to especially hybrid intelligent systems and strong Deep Learning models, we often hear about sensational applications of Artificial Intelligence in healthcare problems. However, there is one tradeoff in which the more advanced intelligent systems we use the more we get solutions beyond of our understanding capabilities. On other words, we may say that today's intelligent systems are more complicated as their inside mechanisms cannot be understood without any additional efforts. So, that causes issues of safety and trust for the interaction between human and the machine. That's even more vital when we consider Artificial Intelligence inside healthcare problems. Although it was able to interpret traditional Machine Learning techniques, it is now not always possible to apply the same thing to more complicated and more successful systems with multiple techniques including even deep analysis capabilities for challenging healthcare issues.

This edited volume *Explainable Machine Learning for Multimedia-Based Healthcare Applications* combines the latest research regarding Explainable Machine Learning, which is among the latest solutions for improving trust and safety level of intelligent systems for healthcare problems. Nowadays, research works are done in the context of multimedia-based data when we consider multiple way of inputs for more effective outcomes of healthcare applications. So, the book focuses more on integrating explainability (and sometimes interpretability aspects) inside the research efforts for medical problems such as diagnosis, tracking, and discoveries. It is also great to see that there are remarkable reviews for present and future potentials of Explainable Artificial Intelligence for healthcare with especially multimedia-based data. I appreciate the general organization of the chapters because with its current form, the book has a wide readership scope from academics to degree students and even general audience. As the book also employs a multidisciplinary approach, the book can be used as a reference work inside the courses of different fields like Computer Science and Engineering, Biomedical, Medical, Mechatronics, Data Science, and the associated engineering areas for software and/or hardware oriented developments. I think the content of the book will give many inspirations for also researchers from beginner to intermediate and advanced levels. I see that the future will be brilliant with a good collaboration between human and the machine, and this book ensures remarkable points to consider for building strong bridges towards healthcare applications. I would like to thank to dear colleagues and editors Dr. Hossain, Dr. Kose, and Dr. Gupta for their great efforts to build such a valuable work for the scientific literature. Also, it is a pleasure for me to be reading from valuable authors around the world, so my special thanks go to them!

Now it is time to prepare a coffee, take a look to the next pages, and learn more about the latest outcomes of Explainable Machine Learning for healthcare applications covering multimedia data. All the best for the new era of human-machine collaboration in healthcare!

National Autonomous
University of Mexico,
Mexico City, Mexico

Jose Antonio Marmolejo-Saucedo

Preface

As a result of many advancements in the field of Artificial Intelligence, it is now easier to process data and derive meaningful outcomes for real-world problems represented in the form of digitally modeled relations. It is clear that especially Deep Learning era brought fast and effective enough intelligent systems, which are capable of dealing with even the most challenging problems we encounter in fields of the modern life. Among these fields, the healthcare has a vital place as it is affecting well-being of individuals and building up sustainable and better future for next generations. Early applications of Artificial Intelligence for healthcare were associated with rule-based expert systems and the last decade of the twentieth century included revolutionary Machine Learning solutions for several problems such as diagnosis, treatment, and healthcare planning. Now, at the time of passing the first quarter of the twenty-first century, Artificial Intelligence has provided already cutting edge solutions for many problems such as cancer diagnosis, advanced treatments, and even drug discovery. Deep Learning keeps the main role in all these advancements so far. On the other hand, it has been even seen that the rise of Internet of Things and wearables made it possible to gather instant data and ensure instant communication cycles for better patient tracking and healthcare management. However, such advanced applications nowadays caused a tradeoff between interpretability and successful performance. Although we are able to obtain great performances with today's Artificial Intelligence-based systems, it is more difficult to understand how such systems processed the input data to derive the successful outcomes. As it may be understood, that's a too critical problem for healthcare applications. Additionally, the most recent applications of hybrid Machine Learning or Deep Learning architectures are able to work on multimedia-based medical data, so it is more important to employ effective mechanisms to make them interpretable for the human side. As a result of advancements through interpretability of Machine Learning, a newer concept of Explainable Artificial Intelligence (XAI) was introduced for the more complex intelligent systems. By ensuring more research on interpretability and explainability of intelligent systems, the future advancements for multimedia-based healthcare applications will be trustworthy and human-compatible.

As based on the current state of the literature, this edited volume provides the most recent research studies regarding Explainable Machine Learning applications for multimedia-based healthcare applications. The content has been a collection of research works aiming to solve different healthcare problem cases and ensuring a wide enough literature review regarding the scope of the book. As a result of meticulous efforts by international authors, we organized a useful reference work, which will be beneficial for researchers, degree students, and even private/public sector experts in the context of healthcare and engineering fields. The most critical contribution of the book is associated with use of multimedia data for healthcare problems and ensuring interpretable and/or explainable intelligent system formations, which are able to ensure effective results for target problems.

In the context of the collection, we reviewed and included a total of 12 chapters briefly as follows:

Chapter "Automatic Fetal Motion Detection from Trajectory of US Videos Based on YOLOv5 and LSTM" reports an automatic fetal motion detection system, which is able to use trajectories of US videos, by employing both YOLOv5 and LSTM models.

Chapter "Explainable Machine Learning (XML) for Multimedia-Based Healthcare Systems: Opportunities, Challenges, Ethical and Future Prospects" provides a comprehensive review regarding Explainable Machine Learning (XML) for multimedia-based healthcare problems. In detail, it also discusses about opportunities, challenges, and even future ways with ethical concerns included.

Chapter "Ensemble Deep Learning Architectures in Bone Cancer Detection Based on Medical Diagnosis in Explainable Artificial Intelligence" comes with an ensemble explainable Deep Learning system formation to provide bone cancer detection.

Chapter "Digital Dermatitis Disease Classification Utilizing Visual Feature Extraction and Various Machine Learning Techniques by Explainable AI" focuses on the problem of digital dermatitis disease classification by building a solution mechanism covering visual feature extractions and Machine Learning usage with Explainable Artificial Intelligence (XAI) aspects.

Chapter "Explainable Machine Learning in Healthcare" reports a good review regarding Explainable Machine Learning for healthcare problems, by focusing on cases with multimedia data.

Chapter "Explainable Artificial Intelligence with Scaling Techniques to Classify Breast Cancer Images" considers the problem of breast cancer and introduces an XAI solution with scaling techniques and classification-based outcomes.

Chapter "A Novel Approach of COVID-19 Estimation Using GIS and Kmeans Clustering: A Case of GEOAI" includes a recent problem: COVID-19, and builds an estimation approach by using GIS and K-means clustering in the context of an interpretable approach.

Chapter "A Brief Review of Explainable Artificial Intelligence Reviews and Methods" is based on a review-based work in which the readers are informed about XAI methodologies and the most recent applications for healthcare problems.

Chapter "Systematic Literature Review in Using Big Data Analytics and XAI Applications in Medical" ensures a systematic literature review for use of Big Data Analytics and XAI applications targeting medical problem cases.

Chapter "Using Explainable Artificial Intelligence in Drug Discovery: A Theoretical Research" provides a review about using XAI for a very critical problem area: Drug Discovery.

Chapter "Application of Interpretable Artificial Intelligence Enabled Cognitive Internet of Things for COVID-19 Pandemics" revisits the problem area of pandemics and introduces an interpretable Artificial Intelligence-enabled Cognitive Internet of Things for the COVID-19 case.

Chapter "Remote Photoplethysmography: Digital Disruption in Health Vital Acquisition" considers a remarkable problem: Remote Photoplethysmography, and reports about the research regarding digital disruption in healthcare scope.

We believe that the content of the book will be giving enough reference information and ideas for building the most recent knowledge and establishing further research studies. We would like to send our warmest regards and congratulations to all chapter authors. Also, our special thanks go to respectful Prof. Jose Antonio Marmolejo-Saucedo (from National Autonomous University of Mexico, Mexico) for his kind Foreword, which improved the value of the book. We would be grateful to receive any ideas, suggestions, and collaborations from all readers around the world. We hope you will enjoy your trip inside the pages of such a valuable, timely work. Thank you!

Riyadh, Saudi Arabia — M. Shamim Hossain
Isparta, Turkey — Utku Kose
Delhi, India — Deepak Gupta

Acknowledgement

As the editors, we would like to thank Shina Harshavardhan, Susan Grove, and the Springer Team for their valuable efforts and great support on pre-organizing the content and publishing of the book.

Contents

Automatic Fetal Motion Detection from Trajectory of US Videos Based on YOLOv5 and LSTM

Musa Turkan, Furkan Ertürk Urfalı, and Emre Dandıl

1 Introduction

Abnormal developments in the fetus, especially in recent years, cause an increase in deaths close to birth for many countries [1]. Almost 48 million babies are stillborn over the past two decades, and average of 2 million babies is stillborn each year, mostly in low-income countries [2]. It is considered that most infant deaths can be prevented with life-saving measures and quality health care. Additionally, 3–6% of babies worldwide are born with a serious defect each year [3]. In addition, maternal mortality due to childbearing and pregnancy is approximately 211 per 100,000 live births [4]. Therefore, detection of an abnormality occurring in the fetus at an early stage is very important for both maternal and fetal health.

In addition to monitoring the development of fetal anatomical structures, some data collected about the fetus can provide strong indications for the development of the fetus. In particular, the examination of fetal images and videos enables early diagnosis of some diseases/disorders. However, recognition and tracking of movements and determination of anatomical structures from fetal video and image scans require difficult and complex processes. Fetal scans are usually performed with

M. Turkan
Institute of Graduate, Department of Computer and Electronics Engineering, Bilecik Seyh Edebali University, Bilecik, Turkey
e-mail: musa.turkan@bilecik.edu.tr

F. E. Urfalı
Medical Faculty, Department of Radiology, Kütahya Health Sciences University, Kütahya, Turkey
e-mail: furkanerturk.urfali@ksbu.edu.tr

E. Dandıl (✉)
Faculty of Engineering, Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik, Turkey
e-mail: emre.dandil@bilecik.edu.tr

M. S. Hossain et al. (eds.), *Explainable Machine Learning for Multimedia Based Healthcare Applications*, https://doi.org/10.1007/978-3-031-38036-5_1

ultrasound (US) imaging. US is widely used in the evaluation of fetal health and development due to its many advantages such as cost, availability, real-time processing [5].

By accurately monitoring the pregnancy process using US data, important anatomical structures in the fetus can be detected and movements can be followed. In addition, the intensity, frequency and duration of fetal movements can be determined by the perception of the movements by the mother. Therefore, automatic detection and evaluation of fetal movements becomes important. In addition, detection of critical health conditions in the fetus, such as the presence of active heartbeat, is associated with fetal movements [6]. In today's routine US scans for the follow-up of fetal health, several biometric evaluations and measurements are performed for the anatomical structure of the fetus, such as measuring the bone lengths of the hands and legs, determining the head circumference and resting the heartbeat. Besides, some examinations such as fetal weight and evaluation of abdominal circumference are used to monitor the health status of the fetus [7]. In addition, the follow-up of fetal movements is an important indicator of a healthy pregnancy process, and decreased or no fetal movements can be considered as a sign that there may be serious risks to the fetus [8]. On the other hand, increased fetal movements may also be a sign of undesirable conditions. Consequently, accurate assessment of complex fetal movements such as this is vital.

It is seen that there are different studies previously suggested for the detection, follow-up and classification of fetal movements. These studies generally include tools and methods that facilitate physicians' decision-making processes. For example, studies to determine various organs and anatomical structures in the fetus [9], to make some biometric measurements such as head circumference and leg length [10], to determine the standard plane in the fetus [11, 12] are some of these studies. In addition, there are studies [13, 14] that provide classification of US images to assist experts in their decision-making. Ishikawa et al. [13] proposed the recognition and classification of fetal parts such as head, trunk, legs and other parts to predict fetal position.

Deep learning algorithms, which can be applied in many fields, have been widely used in the processing of medical data in recent years. In addition, deep learning algorithms continue to improve day by day with higher performance, less hardware consumption and more accurate results. In deep learning models developed especially for the classification of movements, objects can be tracked with their motion trajectories. In this case, the changing position of the moving object in each frame can provide information about the movement. In this way, there are studies that are combined with deep learning methods and that classify using motion trajectories [15–17]. In addition, human movements can be defined by adding the motion trajectories of the joints as a data to the 2D image [18]. Moreover, trajectory information of objects can be used to represent temporal information of motion [19, 20].

In recent years, many studies [21–23] in which deep learning methods are widely used for the evaluation of fetal ultrasound data are gaining importance. van den Heuvel et al. [23] proposed a deep learning method that allows measuring head circumference using ultrasound images in countries where resources may be limited. In another study, Dozen et al. [24] segmented small and dynamically shaped heart structures with a deep learning-based network in fetal cardiac ultrasound videos using time series information and special region information. In their study, Ravishankar et al. [25] provided the detection and measurement of the abdomen region from 2D ultrasound images, using a method that combines traditional tissue recognition methods and deep learning methods. Yaqup et al. [26] proposed a deep learning method for the detection of brain abnormalities in the possible fetus to ensure the correct localization of the fetal brain, the detection of the region of interest, and the recognition of the acoustic pattern in the regions that allow to confirm the plane. Accurate acquisition of fetal planes on ultrasound images is an important consideration for accurate biometric measurements and the application of various diagnoses. In their work, Chen et al. [27] developed a framework that enables standard plane recognition from ultrasound videos. In another study, Gustavo et al. [28] presented a method that detects fetal anatomical structures on ultrasound images and can perform biometric measurements. In their study, Chen et al. [29] proposed a deep learning-based framework that enables the measurement of heart ventricles from 2D ultrasound images of the fetus. In another study, Arnaout et al. [30] developed a neural network method that can identify complex heart conditions before birth using ultrasound images to detect congenital heart conditions.

It is known that trajectory information of fetal movements creates different pattern patterns according to anatomical structures. Therefore, the classification of different movement types can be achieved by obtaining the movement trajectories of the anatomical structures in the fetus. In this chapter, a deep learning approach based on YOLOv5 and LSTM methods is proposed for the detection and recognition of fetal anatomical structures using motion trajectories. In the study, first of all, a dataset is prepared from US videos containing the movements of anatomical structures in the fetus. In this dataset, class and location information are obtained by using the YOLOv5 network, allowing the recognition of fetal anatomical structures (organs). In the next step, using the patterns of the 2D trajectories created by the movements, the anatomical structure of the fetal movement is determined with LSTM deep neural networks. Thus, it is possible to classify fetal movements from orbital images.

The next sections of the chapter are organized as follows. In the second part, the prepared US dataset and the proposed deep learning methods are presented in detail. In the third part, the research results obtained with the experimental studies are given. In the last part, the conclusions obtained are discussed, the experimental results are evaluated and the planned future activities are mentioned.

2 Material and Method

Basically, it can be applied to the network by obtaining a single location point belonging to the object, or it can be applied to a deep learning network by obtaining more than one point to represent the object. Thus, the images obtained from fetal videos can be trained with an object recognition network that enables the recognition of each anatomical structure, and classification of anatomical structures in the fetus can be achieved.

Any object recognition algorithm that draws a bounding box around the object can be used to obtain location information of fetal organs (anatomical structures). For this, first of all, ultrasound images can be labeled in accordance with the infrastructure of the object recognition algorithm used to classify anatomical structures. Software that performs labeling in accordance with the object recognition algorithm can be used for this process. After labeling, the point of the bounding box of the object recognition algorithm is determined as position information, and the position of the anatomical structure is taken in each video frame. Thus, an ordered list of points consisting of x and y coordinates is obtained, which sequentially follows the change of organ movements on the screen. In addition, the position information of the anatomical structure of the fetus may occur as points on the 2D plane, depending on the screen resolution. Different screen resolutions and different position of the organ on the screen can cause these values to be quite different. To eliminate this problem, before using the raw location information, this location information is normalized to be in the range of 0–1. Normalized data is treated as an ordered array of points showing the position of the organ in each frame. A suitable deep learning algorithm, LSTM, can be used to process such sequential data. By processing sequential data with LSTM, the succession of points can be estimated. This ordering process differs for each organ. Thus, by training an LSTM network for differences in the ordering of position information, classification of anatomical structures in the fetus can be achieved.

In this chapter, a hybrid deep learning model based on YOLOv5 and LSTM methods is proposed for the detection and recognition of fetal anatomical structures from motion trajectories on US scans. In the prepared dataset, first of all, the recognition of anatomical structures and their locations are provided by the YOLOv5 algorithm. Then, by using the object recognition algorithm, the position information of the object detected throughout the video is created with the coordinate information in the 2D plane, and the pattern of the 2D trajectories of the movements is created. In the last stage, object motion positions are detected in each frame by using the point data array of the motion trajectories obtained, and the obtained motion points are recognized by the LSTM network. For this study, the methodology of the proposed system for the detection and recognition of fetal anatomical structures from the motion trajectories is shown in Fig. 1.

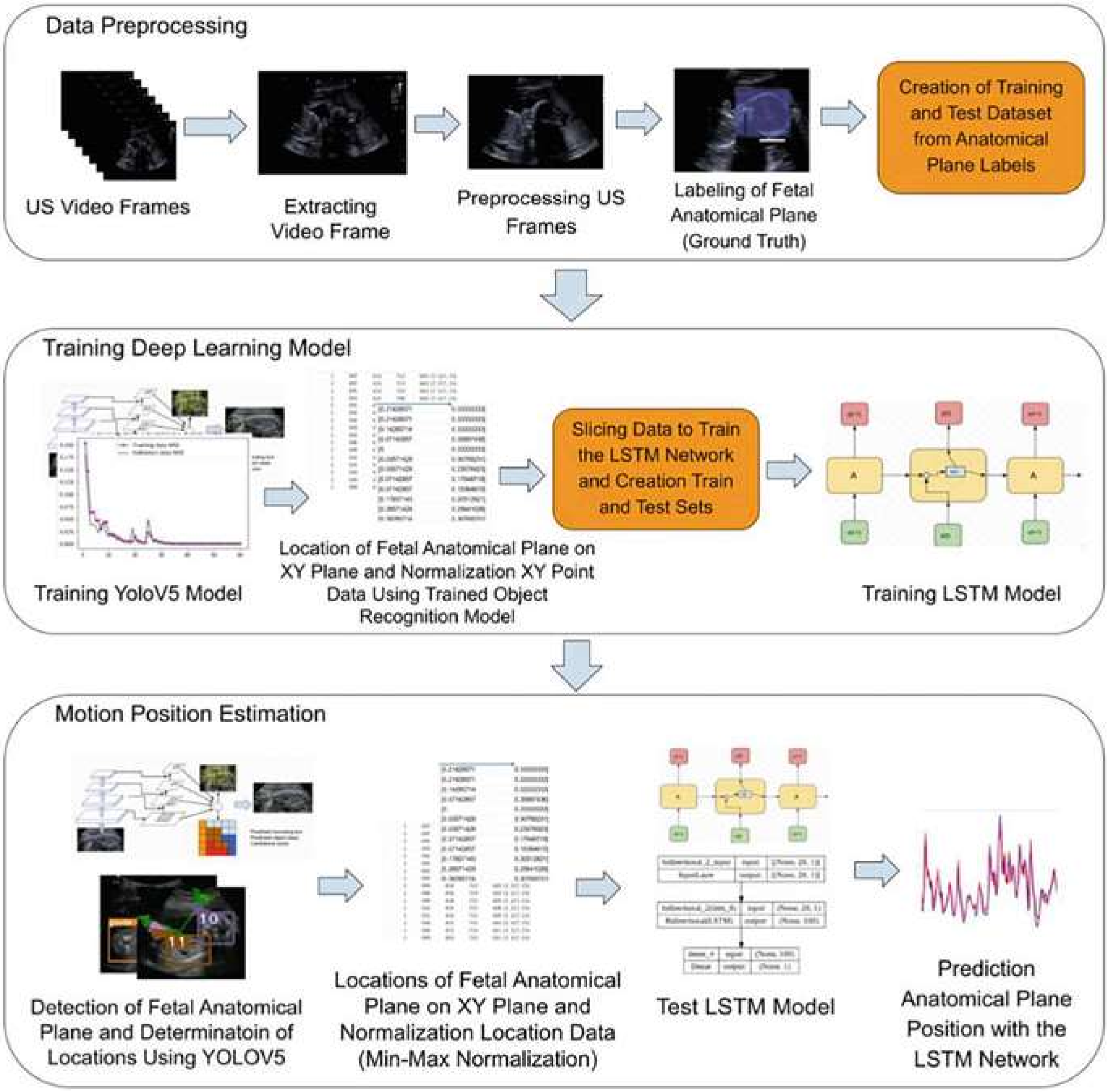


Fig. 1 The methodology of the proposed system for the detection and recognition of fetal anatomical structures from the motion trajectories

2.1 Dataset

In this study, fetal scans used in the prepared dataset for the prediction of fetus' anatomical structures were collected from Evliya Çelebi Training and Research Hospital of Kütahya Health Sciences University. In addition, with the decision of the Ethics Committee of Non-Interventional Clinical Researches of Kütahya Health Sciences University, dated 08.07.2021 and numbered 2021/12–07, it has been confirmed that there is no ethical and scientific inconvenience in the conduct of the study. Data on ultrasound scans were obtained from 10 different volunteers who were pregnant between 16 and 20 weeks. In the dataset, all pregnant women were selected from the second trimester cases because they allow easy evaluation of fetal movements. For all fetus, major extremity movements, diaphragm and swallowing movements and head and body movements were recorded as US video. US

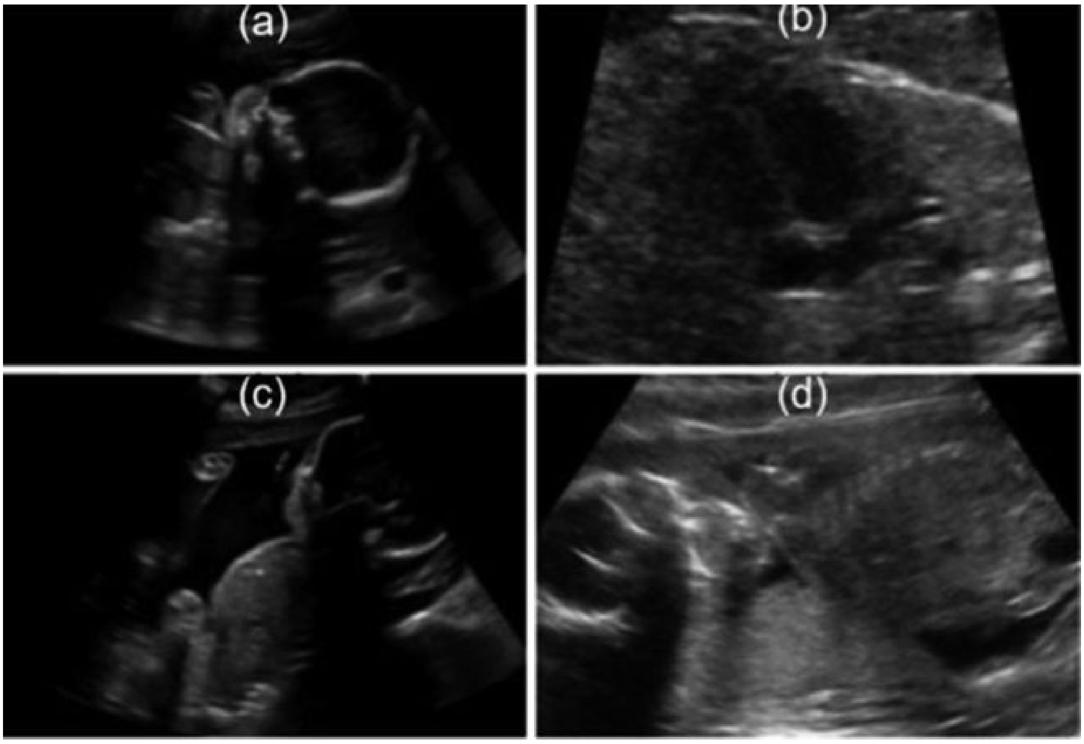


Fig. 2 Frame slices of some anatomical structures in the prepared fetal US dataset (**a**) head, (**b**) heart, (**c**) head and body, and (**d**) head and body

evaluations in fetus were performed using Acuson S3000; Siemens Medical Solutions, Mountain View, CA. This US system includes a convex transducer with 4 MHz frequency and obstetric US preset. Unlike our previous study [5], a total of 2500 2D ultrasound images from US scans were acquired. These images were labeled with the consensus of three different experts and ground truths were created. Motion trajectories of fetal anatomical structures such as body, head, arm, and heart were extracted from fetal US videos in the dataset. Sample frames of the anatomical structures of the fetus from the dataset in this study are shown in Fig. 2.

2.2 *Structure of YOLO v5*

YOLO (you only look once) is a deep learning method proposed by Redmon et al. [31] in 2016 and used in object recognition problems with a bounding box. YOLO is also a one-step algorithm, and with YOLO, both location information and classification of the object can be obtained [32]. Other CNN-based algorithms with accuracy close to the YOLO algorithm require more computational load as they do object class determination and object location determination with separate networks. The created neural networks are applied separately to the image at multiple locations and scales. In the YOLO algorithm, the entire image is passed through the network once.

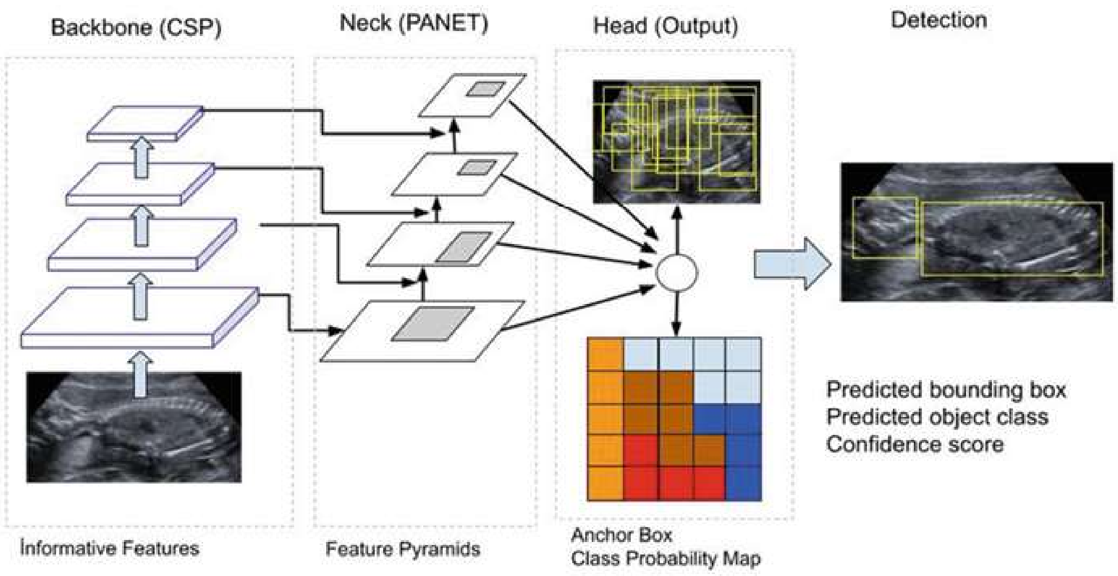


Fig. 3 The modular architecture of the YOLOv5 deep learning model

The YOLOv5 network model was developed by Glenn Jocher [33] in 2020 based on PyTorch. The YOLOv5 model can provide higher speed than previous YOLO versions. Also, the YOLOv5 network generates model files of smaller size and generally takes less time to train. Created with the support of many open source developers, YOLOv5 is based on PyTorch and has different configurations for object recognition from previous versions [34].

The architecture of the YOLOv5 deep learning model, which is a typical one-stage object recognition method, is shown in Fig. 3. This model consists of three modular sub-components: backbone, neck and head [35]. In YOLOv5, the features from the input image are extracted using backbone. In addition, DarkNet framework-based Cross Stage Partial Networks (CSPNet) comes into play as the backbone structure in the YOLOv5 model [36]. In the second module, the neck component, of YOLOv5, the feature pyramid network (FPN) is generated by this component. Feature pyramids are used to recognize the same object in different sizes and resolutions, and in YOLOv5, path aggregation network (PANet) neck is generally preferred to obtain pyramid features. In the last component, head, object recognition is performed as in YOLOv4. Here, the probabilities of the object classes and the bounding boxes are also generated. YOLOv5 includes architecturally different models, and in this study, YOLOv5s, the smallest version, was used to identify fetal anatomical structures.

2.3 *LSTM (Long-Short Term Memory) Deep Neural Networks*

Recurrent neural networks (RNN) are deep learning networks that are often used to predict the next step. Unlike other networks, they can remember, and in RNN networks, the inputs correlate with each other to predict the next step. During the RNN training phase, all relationships between the inputs are remembered and each result feeds the next step [37]. While CNN uses more spatially related data, RNN deep learning structures are models developed for processing time series and sequence data [38] and are widely used in applications such as voice applications, text recognition, author recognition, natural language processing [39].

An RNN cell is shown in Fig. 4. Because the RNN itself contains loops that make the information permanent, A, a network segment, looks at an x(t) input and generates a y(t) output. Thus, a loop allows information to pass from one step of the network to the next.

RNN deep learning models can be thought of as multiple copies of the same network, with each cell transmitting a message to the next. An activation function is implemented in each layer here [40]. Thus, a chain-like structure is formed for the RNN as in Fig. 5. The chain-like structure reveals that RNN is closely related to

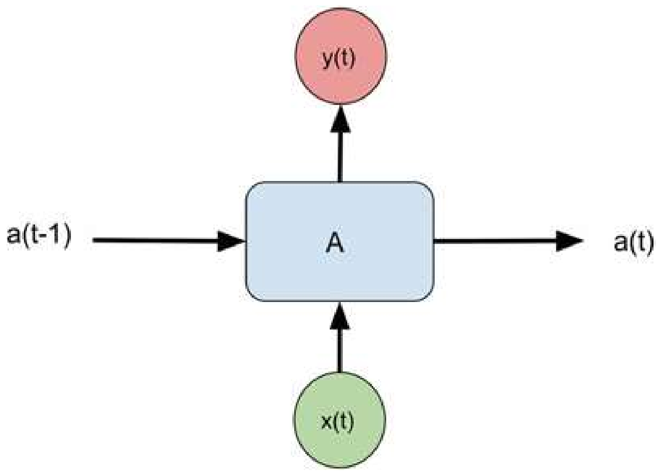


Fig. 4 Recurrent neural network cell structure

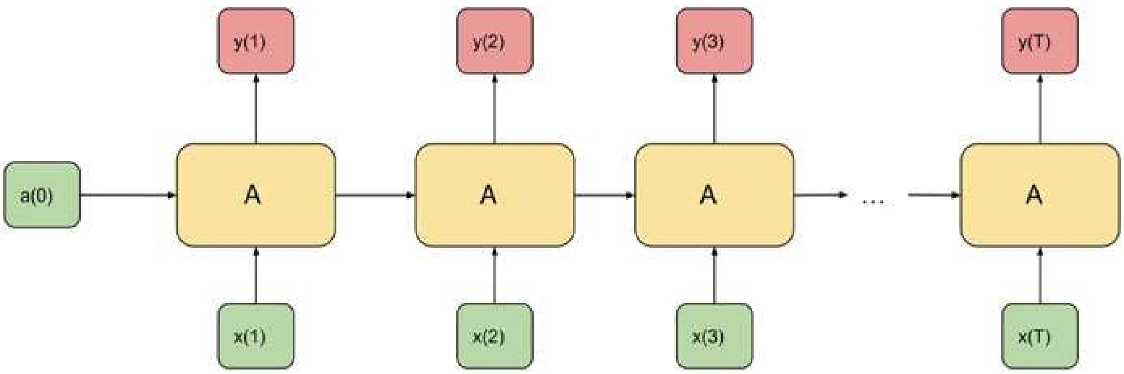


Fig. 5 Layered chain structure of cells in RNN

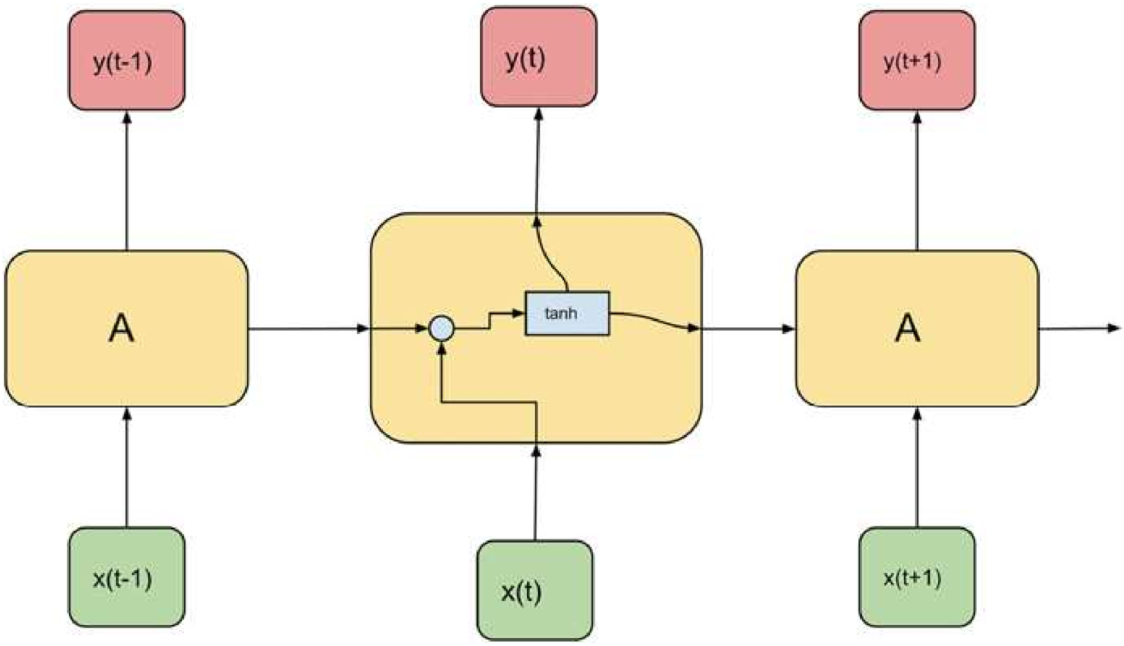


Fig. 6 A single layer of modules in the RNN structure

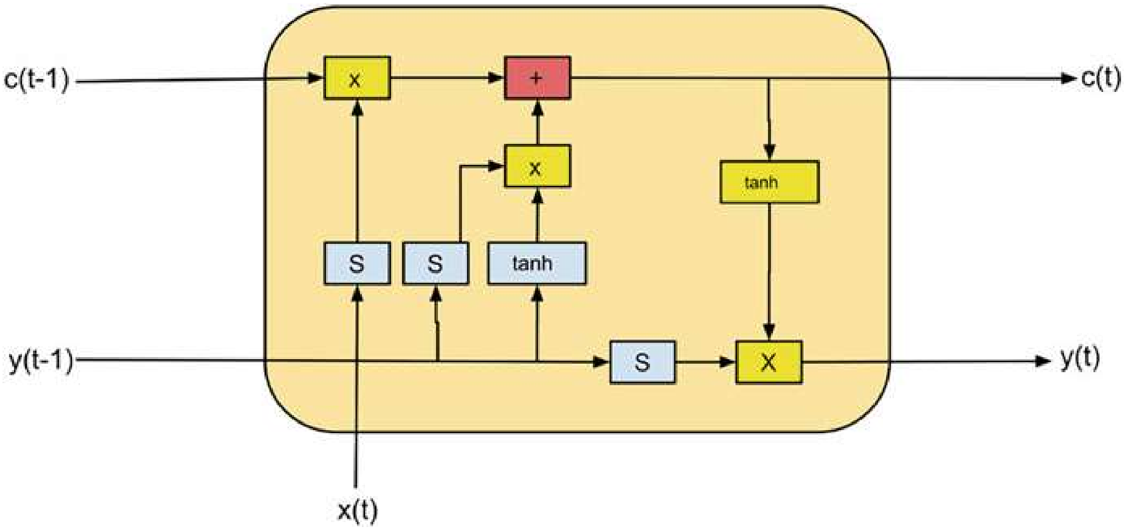


Fig. 7 Cell structure in LSTM deep learning networks

arrays and lists, and is the natural architecture of the neural network to be used for such data. In addition, as seen in Fig. 6, each module in the RNN structure contains a single layer.

The method that makes RNN successful is the LSTM network. Gradient disappearance problem in simple RNN models is solved with LSTM networks [41]. LSTM networks remember the important data and forget the unimportant ones and solve the gradient disappearance problem [42]. Within each repeating module in the LSTM network, there are four layers: input gate, output gate, forget gate, and memory cells [43]. Each layer uses different weight values and the output is calculated using the sigmoid function and the previous state [44]. In Fig. 7, the structure of an LSTM cell module is denoted. Here, tanh is the tangent activation function and S is the sigmoid activation function. In addition, X is the multiplication, + addition, y (t) output, c (t−1) the previous memory cell state information, c (t) the memory state information going to the next cell / cell state, y (t−1) represents the output of the previous cell and x (t) represents the current data input.

3 Experimental Analysis

In this study, anatomical structures in ultrasound images are detected with a hybrid method based on deep learning. First of all, in a fetal US dataset prepared specifically for this study, the positions of anatomical structures in the 2D plane are obtained using YOLOv5. Then, this obtained position information is applied to an LSTM network to predict anatomical structure movements in 2D plane. The results obtained within the scope of the experimental analyze show that the anatomical structures in the fetus are detected successfully with the proposed hybrid method by using the trajectories.

US data for experimental studies were obtained from 10 different volunteers who were pregnant between 16 and 20 weeks. A total of 2500 2D ultrasound images (frames) from US scans were acquired for recognizing the anatomical structure in the fetus using YOLOv5 and obtaining location information. Two thousand of these fetus images are for training, and five hundred images are for test. These frames were delineated with the consensus of three different experts from Evliya Çelebi Training and Research Hospital of Kütahya Health Sciences University, as seen in Fig. 8. LabelImg [45] software tool was used to label US frames by experts.

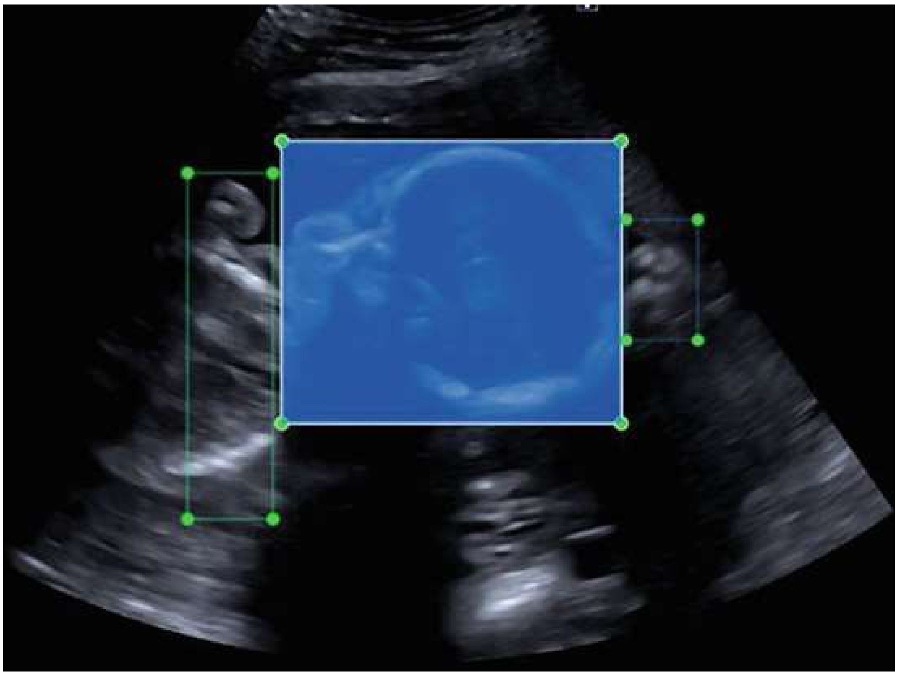


Fig. 8 Labeling of anatomical structures by experts in fetal US images

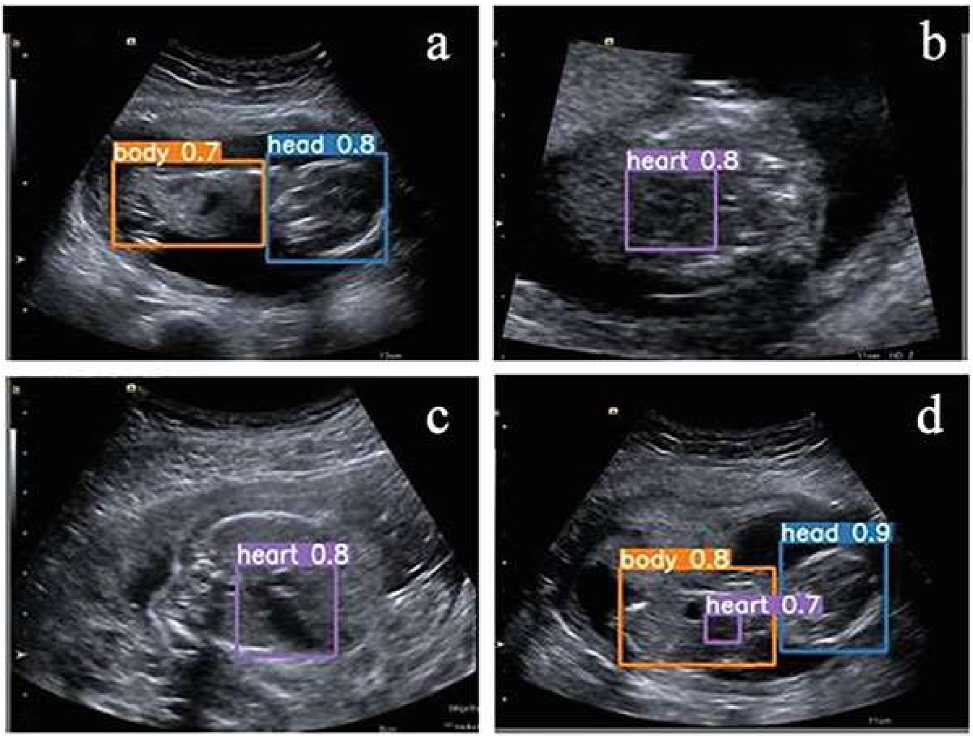


Fig. 9 Recognition of fetal anatomical structures from US videos with YOLOv5 and generating of bounding boxes with detection scores. (**a**) body and head, (**b**) heart, (**c**) heart, and (**d**) body, heart and head

Figure 9 shows the results of the recognition of some anatomical structures such as head, body and heart of the fetus using the YOLOv5 method in the fetal US dataset. From here, it is seen that the anatomical structures of the fetus are recognized successfully with the YOLOv5 based method, and the object is labeled with bounding boxes and its scores are also indicated. In Fig. 9a, the anatomical structures of the body and head are recognized successfully. In addition, the anatomical structure of the heart in Fig. 9b, c and the anatomical structures of the body, head and heart in Fig. 9d are recognized in US videos.

In this study, as in our previous study [5], objects were identified on US videos using the Deep-SORT algorithm. Moreover, using YOLOv5, anatomical structures (like head, heart and body) were recognized from US scans. In addition, their motion trajectories were extracted by tracking their movements. In this regard, Fig. 10 shows the tracking of the motions of anatomical structures identified in the fetal ultrasound with the Deep-SORT algorithm during the US video and the extraction of motion trajectories. Here, the upper left corner of the bounding box of the fetal anatomical structure was used as the base for the position information. Moreover, the trajectory of the detected anatomical planes was formed and 2D trajectory patterns were created with different colors.

After extracting the trajectories of the anatomical structures on the fetal images, the motion trajectories of the anatomical structures were obtained as raw data as csv files. The 2D signals created for the heart and head with motion trajectories are

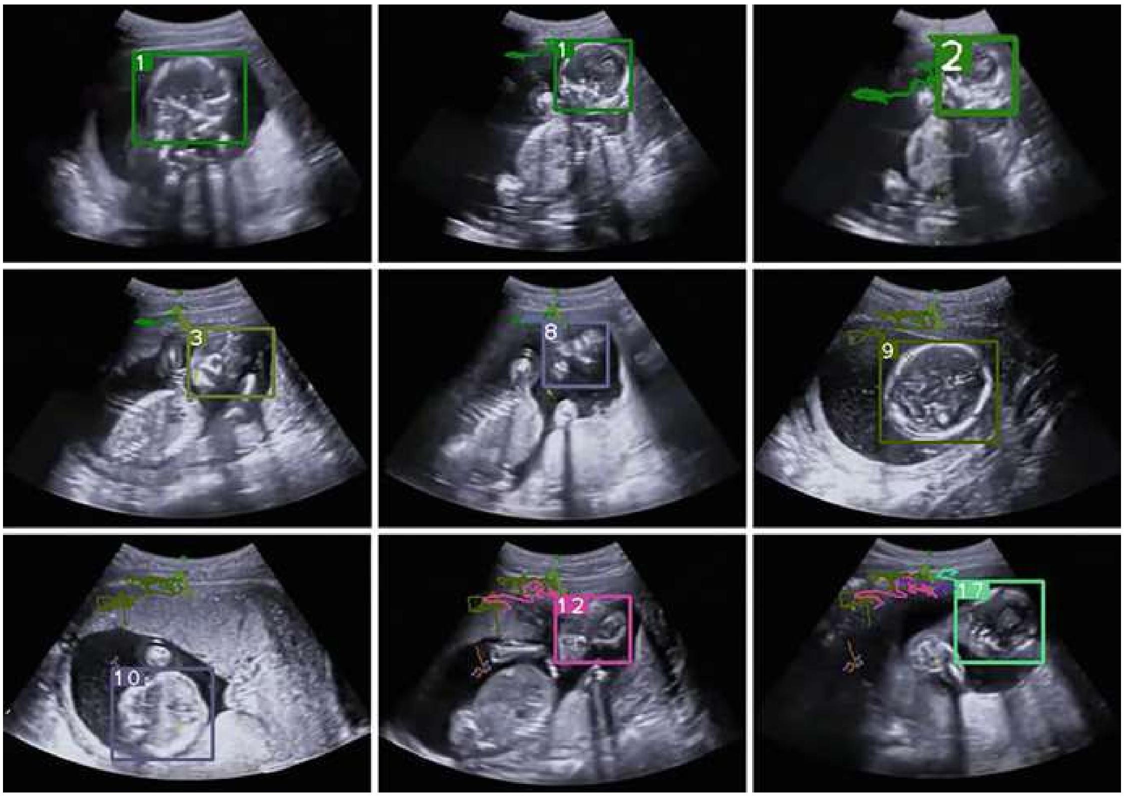


Fig. 10 Extracting the patterns of the motion trajectories of the anatomical structures of the fetus

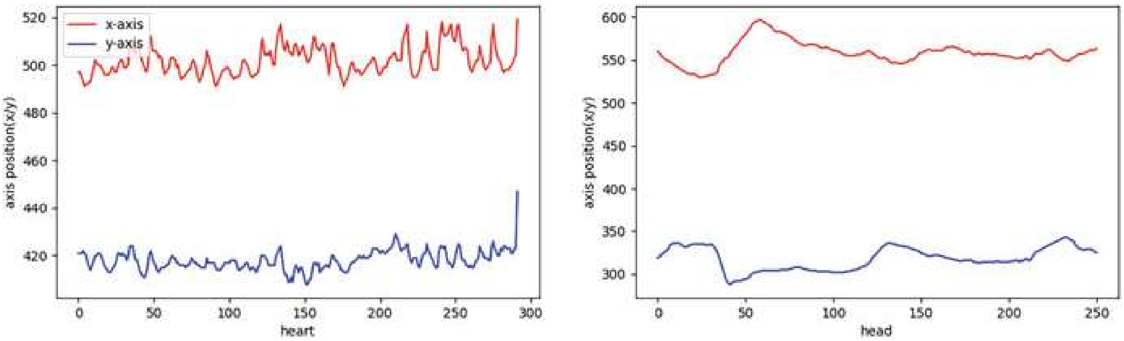


Fig. 11 2D signals created for heart and head with motion trajectories

presented in Fig. 11. In addition, different values may appear in the trajectory data obtained depending on the width and height of the screen. Depending on where the anatomical structure of the fetus is located on the screen, x and y coordinate information may also change. For this reason, in order to make the trajectory data more meaningful, the obtained signals were subjected to min-max normalization.

When the movement points of the anatomical structures in the fetus are combined with a line on the screen, different patterns emerge. Therefore, differences in movements can be detected when the position information of the anatomical structures in the fetus is displayed in a 2D plane as a graph showing the x and y points, as in Fig. 11. From this point of view, a corner of the bounding box determined by the

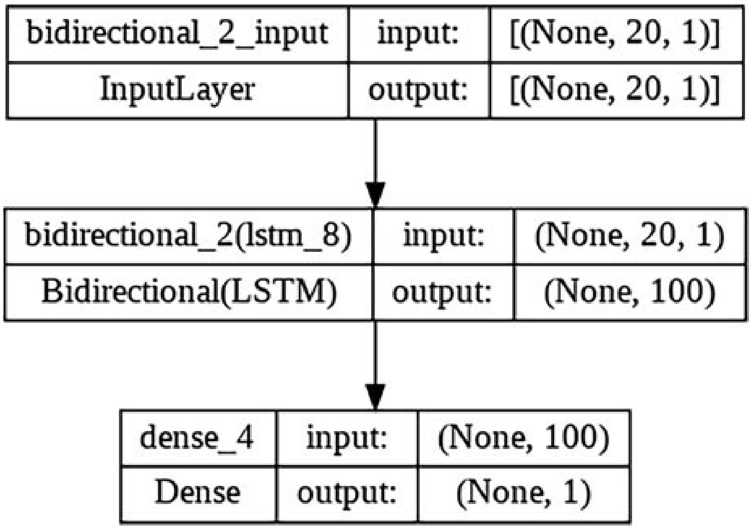


Fig. 12 Structure of LSTM network used for classification of trajectory data converted into 2D signal array

algorithm or the midpoint of the bounding box can be used when determining the position information of the anatomical structures to be referenced. In some movements, the reference to the midpoint may not make a descriptive difference, but the midpoint of the object will not change for movements such as heartbeat movement. For such cases, it would be more appropriate to consider the corner points of the object. Instead of a single vertex of the object, all vertices of the bounding box can be obtained as trajectory data. In this case, additional features such as the changing size of the object on the screen can also be used for classification purposes.

When the position information is taken by reference to the same point for each anatomical structure in the US image, a dataset in the form of a sequence is formed. Trajectory data can be examined by obtaining methods such as evaluating each point as a vector with x and y components, examining the change of only one axis. The LSTM method, which is used to predict the next data in a series of data, is a suitable method for examining the sequential trajectory data depending on time. Therefore, in this study, trajectory data converted into 2D signal sequences were classified with LSTM and it was determined from which anatomical structure the movement in the fetus originated. This trajectory information was collected as much as possible from different videos according to the anatomical structure type and a dataset was created for training the LSTM network.

The LSTM network works with arrays with a certain number of elements as input. When dividing the position data into sequences, it is necessary to use the successive position information to reveal the differences of the movement. The sequential position data obtained according to the number of elements of each array to be applied to the LSTM network has to be divided. Therefore, a sequence of points sufficient to define the motion in accordance with the LSTM network for training was determined and applied as an input to the LSTM network. In this study, the structure of the LSTM network used for classification of trajectory data converted into 2D signal array is shown in Fig. 12. Training and testing are done using a fully connected layer at the output of the LSTM network. To train this LSTM network, the action points are divided into arrays of 20 points. One hundred movement point data

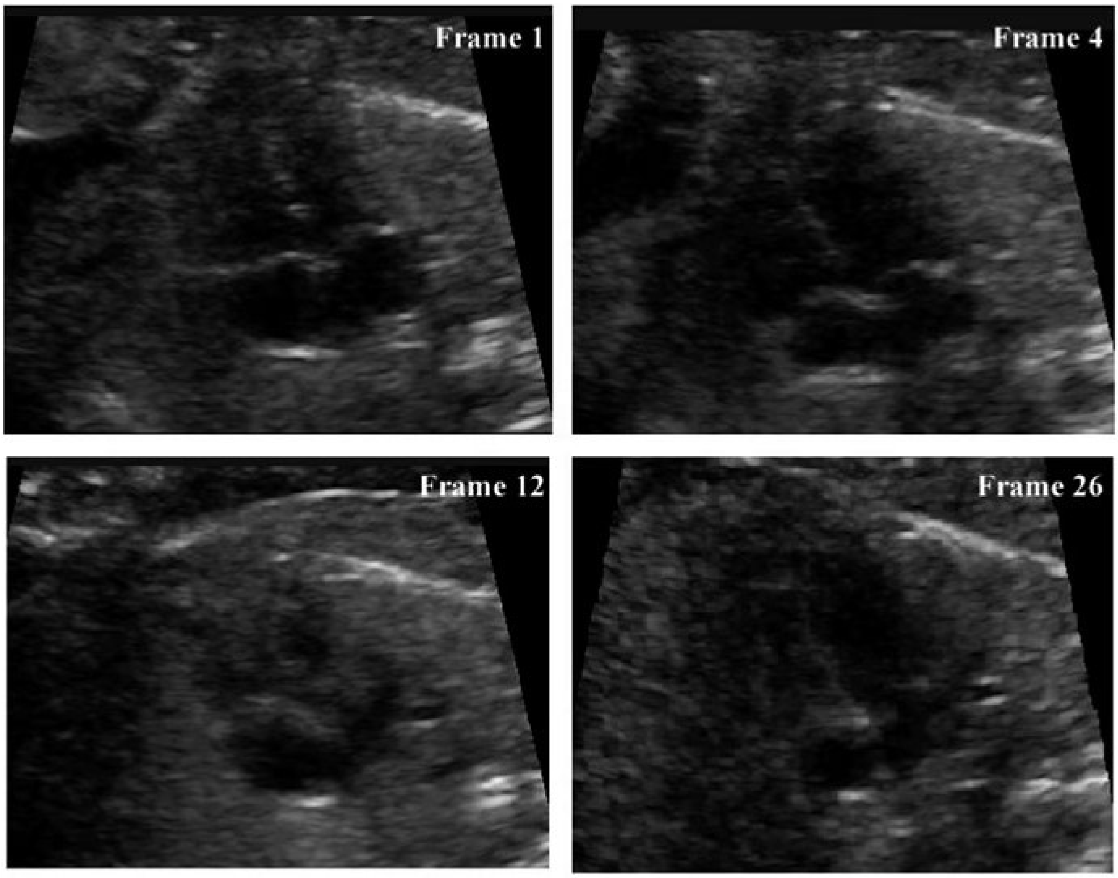


Fig. 13 Changes in the movements of the anatomical structure of fetal heart in different US frames

that was not used in the training was also used in the testing processes. The created LSTM network is trained for 100 epochs. In addition, the motion trajectory data of the anatomical structures of the fetus were created on a computer equipped with an Intel Core i5 CPU and an Intel UHD Graphics 620 graphics card. The normalization of the data, training and testing of the LSTM network was performed on a computer with a 2.20 GHz Xeon(R) CPU and an NVIDIA Tesla T4 GPU via Google Colab.

In Fig. 13, four frames of fetal heart anatomical structure obtained from US videos and fetal movements are denoted. In Frame 1, Frame 4, Frame 12 and Frame 26, it is seen that the visual pattern of the heart and therefore its movements have changed. From here, the anatomical structure can be classified by extracting the pattern of the motion trajectory in each frame in the US videos of the heart.

As seen in Fig. 14a, when the movement data of the heart recognized by YOLOv5 in the fetus are normalized, a graph is obtained with the x and y components changing as in Fig. 14b. The size of the heart changes during heartbeats, and therefore the position of the bounding box may also change. This change can enable heartbeat movements to be captured. In experimental studies with reference to the midpoint of the bounding box, it may not be possible to capture the differences in motion where the size of the object changes, such as heartbeat motion. Changes in

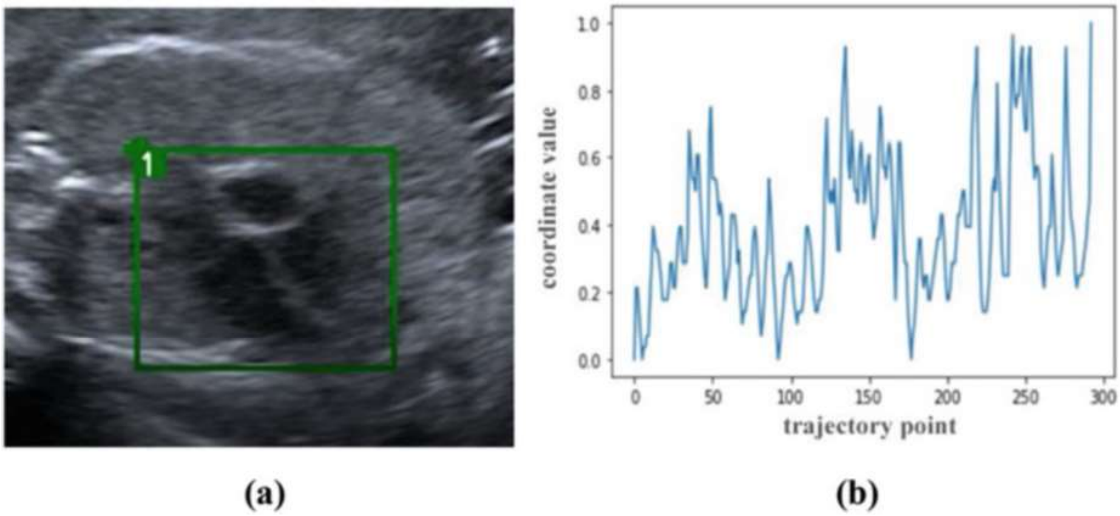


Fig. 14 (**a**) Recognizing the anatomical structure of the heart from the video image of the fetus using YOLOv5 and generating the bounding box, (**b**) Normalizing the motion trajectory position data for the x-axis of the heart

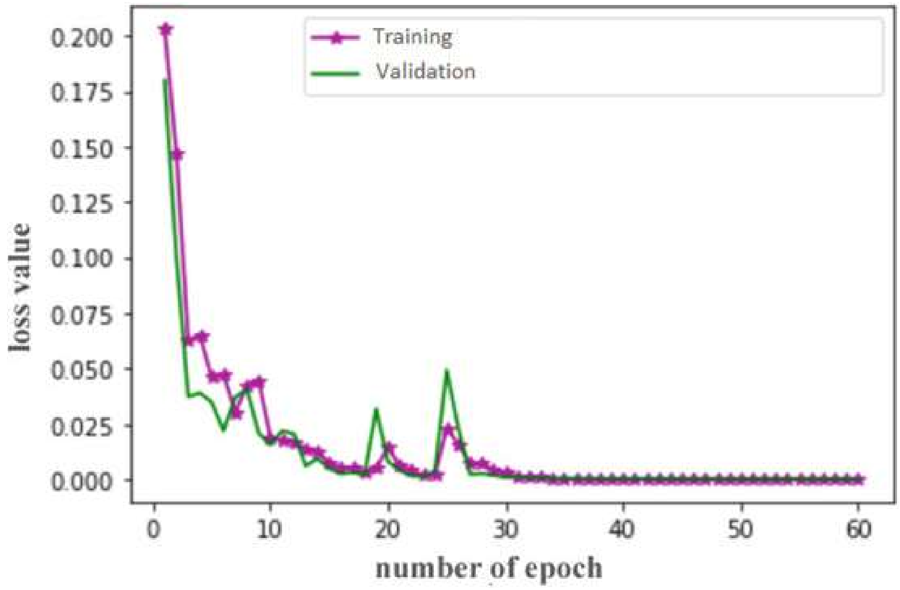
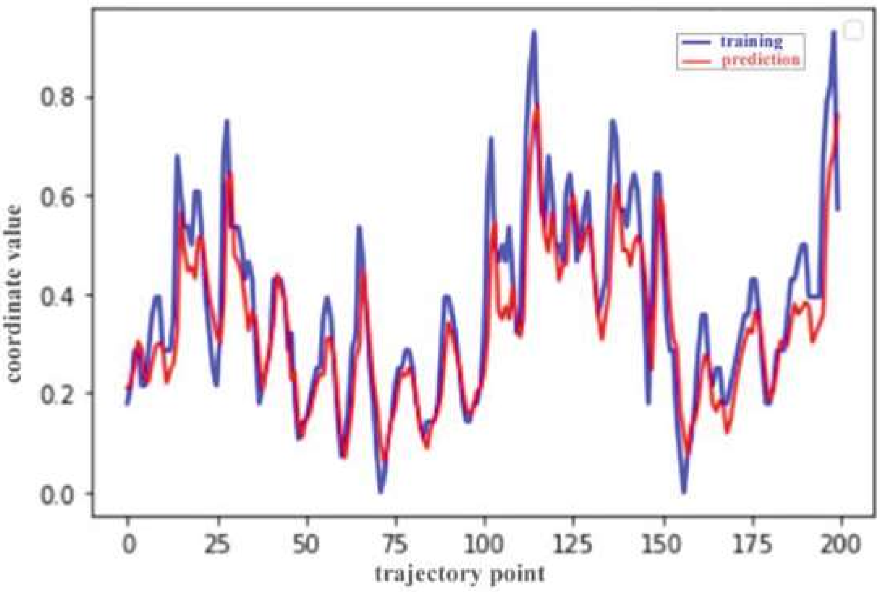


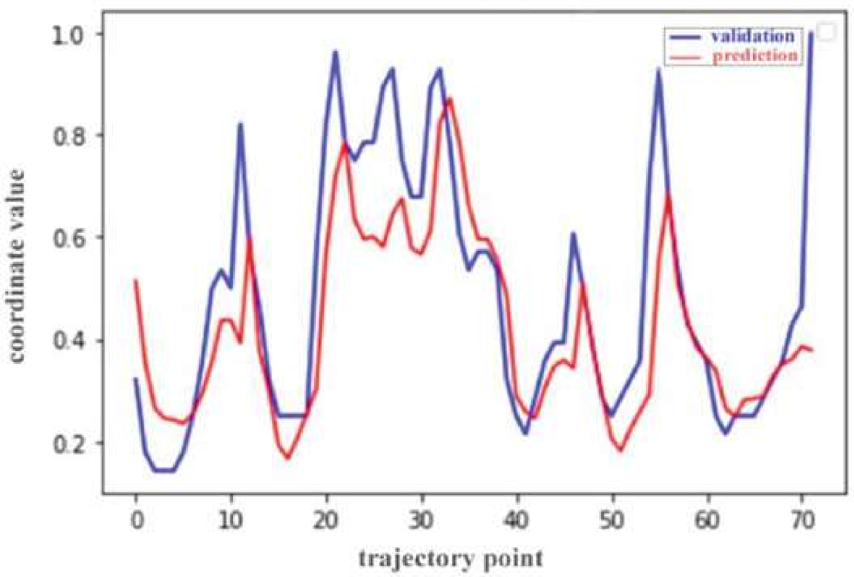
Fig. 15 Loss function change in LSTM network training and validation

heart rate motion can be observed more accurately if a point is referenced from the outer edges of the bounding box.

Approximately 21 thousand parameters are trained in the LSTM network created for the estimation of anatomical structures from 2D signal data generated from motion trajectories. In this LSTM network, there is a bi-directional LSTM (Bi-LSTM) in the first layer, an intermediate layer consisting of 50 cells, and an output layer with a single output. ReLu is used as the activation function in the middleware and the output is estimated with a softmax function. The position data is split into 20 elements and applied to the LSTM network. The graph of loss function change in LSTM network training is as seen in Fig. 15.



(a)



(b)

Fig. 16 (**a**) Prediction of anatomical structure from fetal heart movements of the proposed LSTM network for training data, (**b**) prediction of anatomical structure from fetal heart movements of the proposed LSTM network for validation data

In experimental studies, fetal anatomical movements were predicted on training and test data using the LSTM network trained with 100 epochs. Figure 16a shows a comparison of training and prediction values for the heart along the x-axis, and Fig. 16b shows a comparison of validation and prediction results in analyzes with test data not included in training. While the mean square error (MSE) value for estimation with training data was 0.0069, the mean MSE value for validation data was calculated as 0.0097. From this, it can be seen that the prediction results generally agree with the real data and the LSTM network can recognize the characteristics of heart movements from both training and validation sets.

4 Conclusion

In this chapter, a deep learning approach is proposed by combining YOLOv5 and LSTM methods for detecting and recognizing fetal anatomical structures such as head, heart, body and arm using motion trajectories. First of all, anatomical structures in the dataset created from US videos were recognized by using the YOLOv5 network. Then, using the signal patterns of the 2D trajectories created by the movements, the anatomical structure of the fetal movement was predicted using LSTM deep neural networks. Thus, classification of fetal movements from trajectory images has been achieved.

In fact, there are many studies using trajectory information for motion classification. However, there are not many studies where objects in ultrasound videos can be classified using trajectory information. In the study, it has been observed that when the fetal movement trajectory information is normalized and trained in the LSTM network, the network can adapt to the movements of the organ, and therefore objects in ultrasound videos can be classified using trajectory information. In addition, it is a difficult process to obtain videos in which anatomical structures move in fetal ultrasound scans and to collect videos with sufficient clarity for each anatomical structure. Also, since the fetus does not always move, it is another challenge to wait for appropriate movements to occur to recognize movements at the time of video acquisition.

Instead of using a single point of the object in fetal videos, more point data of movement can be collected by using other points of the bounding box. Using these data can increase the classification performance of the movement. Using this feature of the network, it is possible to predict abnormal motion patterns. Obtaining the movements of fetuses with anomalies is a difficult procedure. Also, disturbances that cause motion anomalies are rare, so collecting such fetal motion videos is also challenging. If ultrasound videos with abnormal movements can be obtained in practice, it is considered that these movements can be predicted. Besides, deep learning models can be used to obtain 3D scans. Generative adversarial network (GAN) is used to create such images, to obtain artificial images that are very lifelike. It may be possible to obtain higher resolution anatomical structure images by applying GAN networks to fetal anatomical structures. Finally, converting the fetus images to high resolution with such a mesh can provide clearer images with ordinary ultrasound devices without the need for tedious operations such as detailed ultrasound.

Acknowledgments We would like to thank Evliya Çelebi Training and Research Hospital of Kütahya Health Sciences University for providing the dataset used in this study and Ethics Committee of Non-Interventional Clinical Researches of Kütahya Health Sciences University for its approval.

References

1. Salomon, L. J., Alfirevic, Z., Berghella, V., Bilardo, C., Hernandez-Andrade, E., Johnsen, S., Kalache, K., Leung, K. Y., Malinger, G., & Munoz, H. (2011). Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. *Ultrasound in Obstetrics & Gynecology, 37*, 116–126.
2. Report of the UN Inter-agency Group for Child Mortality Estimation, 2020-A Neglected Tragedy The global burden of stillbirths, 2020. Accessed 30 Oct 2022. https://childmortality.org/wp-content/uploads/2020/10/UN-IGME-2020-Stillbirth-Report.pdf
3. World Birth Defects Day 2022: Global efforts to prevent birth defects and support families, 2022. Accessed 30 Oct 2022. https://www.cdc.gov/globalhealth/stories/2022/world-birth-defects-day-2022.html
4. Trends in Maternal Mortality 2000 to 2017 Estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division, 2019. Accessed 29 Oct 2022. https://apps.who.int/iris/bitstream/handle/10665/327596/WHO-RHR-19.23-eng.pdf?sequence=13&isAllowed=y
5. Dandıl, E., Turkan, M., Urfalı, F. E., Bıyık, İ., & Korkmaz, M. (2021). Fetal movement detection and anatomical plane recognition using YOLOv5 network in ultrasound scans. *Avrupa Bilim ve Teknoloji Dergisi, 26*, 208–216.
6. Wróbel, J., Kupka, T., Horoba, K., Matonia, A., Roj, D., & Jeżewski, J. (2014). Automated detection of fetal movements in Doppler ultrasound signals versus maternal perception. *Journal of Medical Informatics & Technologies, 23*, 43.
7. Deepika, P., Suresh, R., & Pabitha, P. (2021). Defending against child death: Deep learning-based diagnosis method for abnormal identification of fetus ultrasound Images. *Computational Intelligence, 37*, 128–154.
8. You, J., Li, Q., Guo, Z., & Zhao, R. (2017). Smart fetal monitoring. In *International conference on information science and applications* (pp. 494–503). Springer.
9. Yaqub, M., Napolitano, R., Ioannou, C., Papageorghiou, A., & Noble, J. A. (2012). Automatic detection of local fetal brain structures in ultrasound images. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)* (pp. 1555–1558). IEEE.
10. Sobhaninia, Z., Rafiei, S., Emami, A., Karimi, N., Najarian, K., Samavi, S., & Soroushmehr, S. R. (2019). Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning. In *2019 41st annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6545–6548). IEEE.
11. Lei, B., Zhuo, L., Chen, S., Li, S., Ni, D., & Wang, T. (2014). Automatic recognition of fetal standard plane in ultrasound image. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)* (pp. 85–88). IEEE.
12. Yu, Z., Ni, D., Chen, S., Li, S., Wang, T., & Lei, B. (2016). Fetal facial standard plane recognition via very deep convolutional networks. In *2016 38th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 627–630). IEEE.
13. Ishikawa, G., Xu, R., Ohya, J., & Iwata, H. (2019). Detecting a fetus in ultrasound images using grad CAM and locating the fetus in the uterus. In *ICPRAM* (pp. 181–189).
14. Malathi, G., & Shanthi, V. (2009). Wavelet based features for ultrasound placenta images classification. In *2009 second international conference on emerging trends in engineering & technology* (pp. 341–345). IEEE.
15. Wang, L., Qiao, Y., & Tang, X. (2015). Action recognition with trajectory-pooled deep-convolutional descriptors. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4305–4314).
16. Roy, P., & Bilodeau, G.-A. (2019). Adversarially learned abnormal trajectory classifier. In *2019 16th Conference on Computer and Robot Vision (CRV)* (pp. 65–72). IEEE.
17. Shi, Y., Zeng, W., Huang, T., & Wang, Y. (2015). Learning deep trajectory descriptor for action recognition in videos using deep neural networks. In *2015 IEEE international conference on multimedia and expo (ICME)* (pp. 1–6). IEEE.

18. Wang, P., Li, Z., Hou, Y., & Li, W. (2016). Action recognition based on joint trajectory maps using convolutional neural networks. In *Proceedings of the 24th ACM international conference on multimedia* (pp. 102–106).
19. Abdul-Azim, H. A., & Hemayed, E. E. (2015). Human action recognition using trajectory-based representation. *Egyptian Informatics Journal, 16*, 187–198.
20. Shi, Y., Tian, Y., Wang, Y., & Huang, T. (2017). Sequential deep trajectory descriptor for action recognition with three-stream CNN. *IEEE Transactions on Multimedia, 19*, 1510–1520.
21. Gao, Y., Maraci, M. A., & Noble, J. A. (2016). Describing ultrasound video content using deep convolutional neural networks. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* (pp. 787–790). IEEE.
22. Sinclair, M., Baumgartner, C. F., Matthew, J., Bai, W., Martinez, J. C., Li, Y., Smith, S., Knight, C. L., Kainz, B., & Hajnal, J. (2018). Human-level performance on automatic head biometrics in fetal ultrasound using fully convolutional neural networks. In *2018 40th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 714–717). IEEE.
23. van den Heuvel, T. L., Petros, H., Santini, S., de Korte, C. L., & van Ginneken, B. (2019). Automated fetal head detection and circumference estimation from free-hand ultrasound sweeps using deep learning in resource-limited countries. *Ultrasound in Medicine Biology, 45*, 773–785.
24. Dozen, A., Komatsu, M., Sakai, A., Komatsu, R., Shozu, K., Machino, H., Yasutomi, S., Arakaki, T., Asada, K., & Kaneko, S. (2020). Image segmentation of the ventricular septum in fetal cardiac ultrasound videos based on deep learning using time-series information. *Biomolecules, 10*, 1526.
25. Ravishankar, H., Prabhu, S. M., Vaidya, V., & Singhal, N. (2016). Hybrid approach for automatic segmentation of fetal abdomen from ultrasound images using deep learning. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* (pp. 779–782). IEEE.
26. Yaqub, M., Kelly, B., Papageorghiou, A. T., & Noble, J. A. (2017). A deep learning solution for automatic fetal neurosonographic diagnostic plane verification using clinical standard constraints. *Ultrasound in Medicine Biology, 43*, 2925–2933.
27. Chen, H., Dou, Q., Ni, D., Cheng, J.-Z., Qin, J., Li, S., & Heng, P.-A. (2015). Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. In *International conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 507–514). Springer.
28. Carneiro, G., Georgescu, B., Good, S., & Comaniciu, D. (2008). Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. *IEEE Transactions on Medical Imaging, 27*, 1342–1355.
29. Chen, X., He, M., Dan, T., Wang, N., Lin, M., Zhang, L., Xian, J., Cai, H., & Xie, H. (2020). Automatic measurements of fetal lateral ventricles in 2d ultrasound images using deep learning. *Frontiers in Neurology, 11*, 526.
30. Arnaout, R., Curran, L., Zhao, Y., Levine, J. C., Chinn, E., & Moon-Grady, A. J. (2021). An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. *Nature Medicine, 27*(5), 882–891.
31. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 779–788).
32. Chen, K., Li, H., Li, C., Zhao, X., Wu, S., Duan, Y., & Wang, J. (2022). An automatic defect detection system for petrochemical pipeline based on Cycle-GAN and YOLO v5. *Sensors, 22*, 7907.
33. YOLOv5. (2020). Accessed 11 Feb 2022. https://github.com/ultralytics/yolov5
34. Malta, A., Mendes, M., & Farinha, T. (2021). Augmented reality maintenance assistant using yolov5. *Applied Sciences, 11*, 4758.

35. Liu, W., Wang, Z., Zhou, B., Yang, S., & Gong, Z. (2021). Real-time signal light detection based on yolov5 for railway. In *IOP conference series: Earth and environmental science* (p. 042069). IOP Publishing.
36. Wang, C.-Y., Liao, H.-Y. M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., & Yeh, I.-H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 390–391).
37. Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM--A tutorial into long short-term memory recurrent neural networks, *arXiv* preprint arXiv:1909.09586.
38. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
39. Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena, 404*, 132306.
40. Lei, C. (2021). RNN. In *Deep learning and practice with MindSpore* (pp. 83–93). Springer.
41. Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., & Xu, W. (2016). Cnn-rnn: A unified framework for multi-label image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2285–2294).
42. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*, 1735–1780.
43. Park, K., Choi, Y., Choi, W. J., Ryu, H.-Y., & Kim, H. (2020). LSTM-based battery remaining useful life prediction with multi-channel charging profiles, Ieee. *Access, 8*, 20786–20798.
44. Williams, G., Baxter, R., He, H., Hawkins, S., & Gu, L. (2002). A comparative study of RNN for outlier detection in data mining. In *2002 IEEE international conference on data mining, 2002. Proceedings* (pp. 709–712). IEEE.
45. Labelimg. (2022). Accessed 16 Aug 2022. https://github.com/heartexlabs/labelImg