Deep Learning Using MobileNet for Personal Recognizing

Iman Askerzade
Ankara University of Turkey and Insitute of Physics Azerbaijan Academy of Sciences

Yilmaz Kaya
Batman Üniversitesi

Deep Learning Using MobileNet for Personal Recognizing

Şafak Kılıç
Department of Computer Engineering
Siirt University
Siirt, Turkey
safakkilic@outlook.com

İman Askerzade
Department of Computer Engineering
Ankara University
Ankara, Turkey
imasker@eng.ankara.edu.tr

Yılmaz Kaya
Department of Computer Engineering
Siirt University
Siirt, Turkey
yilmazkaya1977@gmail.com

***Abstract*— The usage areas of biometric technologies are increasing day by day. As the importance of information security becomes more important for people every day, one of the most used areas has been information security. In recent years, human-computer interactive systems have started to attract academic and commercial interest and it is aimed to solve problems such as person recognition, gender estimation, age estimation with these systems. In our study, person recognition was performed through the data collected using wearable sensors. The Daily and Sports Activities data set, which we obtained from the UCI database, has been tested with the developed MobilNet architecture. It has been seen that the data obtained from the sensors are successful in the person recognition problem. The developed system has realized the person identification with 19 different physical movements and also provided the detection of 19 different movements. In addition, success rates were obtained according to the region where the sensors were installed. Thanks to the results obtained in this study, it has been seen that accelerometer, gyroscope and magnetometric sensors are successful in biometric person recognition. In summary, it has been determined that the proposed method is successful in biometric person recognition, thanks to the data obtained from wearable sensors.**

Keywords—Transfer Deep Learning Models, MobileNet, Person Identification, Wearable Sensor, Biometric System

I. Introduction

In the past few decades, the problem of identifying people has been one of the hot areas where researchers emphasize the use of various methods. The human body has several unique characteristics. Some systems can detect these characteristics and distinguish them from others. A system that acknowledges people supported their physical or behavioural characteristics is termed a biometric system. Personal biometric authentication includes distinctive people supported their physiological and/or behavioural characteristics [1]. Biometry technology checks the physical or behavioural characteristics that a private will acknowledge. The biometric system works in 2 ways: (1) identification (also referred to as "identity verification") and (2) authentication (also referred to as "identity verification"). First, the verification of private identity is completed by finding a match within the info of everybody within the information (one-to-many comparison strategy). within the latter, a human biometric info is compared with its example hold on within the system information to verify a human identity.

Physical and behavioral features are the two basic qualities of persons. Physical properties are those that are stable and do not vary over time, whereas behavioral properties are those that change over time and in response to environmental factors. Biometric recognition systems are generally developed thanks to these two characteristics of people. While biometric systems created with data such as facial recognition, fingerprint recognition, hand goometry, iris and retina data are systems created by physiological features, biometric systems created with data such as walking, signature and speech are based on behavioral features. These strategies, on the other hand, have a significant disadvantage in that they can be duplicated [2; 3; 4]

Mimic sounds, the use of duplicate irises, and disguised glasses can be examples of these scams. As a result, new descriptive systems based on individual behavior or features, known as biometrics, based on signals measured from various parts of the body, have been adopted in recent years [5; 6]. Different medical signals are also employed as biometric data, according to studies. Biometric systems are developed using EEG and other signals [7; 8; 6; 9], electrocardiogram [10; 11; 12; 13; 14; 15], and accelerometer [16; 17]. Medical indicators are unique to each person, according to studies [8; 18; 10].

In the study of Alyasseri et al. [19], people are recognized using multichannel EEG waves. In addition, active EEG channels were uncovered by the researchers. The process of recognizing persons is done by the use of electrical signals in the brain, according to Sun et alresearch .'s [9]. They discovered that applying the conventional 1D-LSTM deep learning algorithm to 16-channel EEG measurements resulted in a success rate of 99.56 percent. In their study of identifying people using EEG data, Rodrigues et al. discovered an 87 percent success rate [6].

The person recognition problem was attempted to be solved by using sensor signals in a study by Kılıç et al. [20; 21]. The sensor signals were converted into pictures with various processes and the success of the system was tested with local binary pattern and deep learning networks and models. In both of his articles, the author showed a success rate of over 95% in the person recognition study from sensor tests.

Although the CNN training strategy from scratch can be successful in many problems, the correct optimization of the hyper-parameters in the architecture to be installed is still a difficult process [34].At the same time, a large amount of data is required for a scratch-training technique [35]. However, it is possible to reach high success rates faster and more precisely by training deep architectures, which are enriched with techniques newly introduced to the literature and also do not need hyperparameter optimization, with transfer learning or fine-tuning strategy [36]. The MobileNet architecture is chosen as the training model in this study because it is learned using deep transfer networks, has a low transaction cost, and is ideal for mobile applications.At the end of the study, higher success rates were achieved with the right education strategy compared to other studies in the literature.

In our study, after this stage, the data set was introduced, the models used were shown, and the experimental results obtained were presented in the discussion section. In the last part, the general achievements obtained at the end of my study are given.

II. Data Set

As a part of this study, dataset which is known as a Daily and Sports Activities obtained from the ICU database [22; 23; 24]. This study used Xsens MTx sensors mounted to the person's assigned location to collect data from the 19 previously indicated behaviors (activities). To collect data, place these sensors in 5 different areas of the topic. The chest, right wrist joint, left wrist, right (above knee), and left leg (above knee) have all been identified as possible locations for the device (Figur1). In each Xsens MTx device, there are nine sensors (accelerometer x, y, z; gyroscope x, y, z and magnetometer x, y, z).

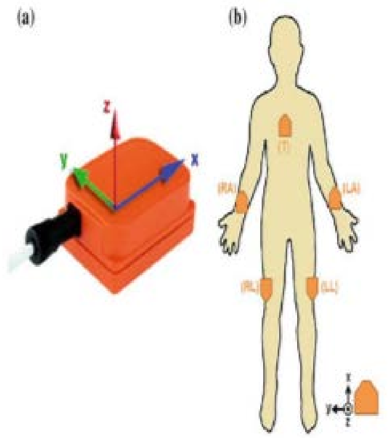

Figure I. Attaching the sensors to 5 different regions of the subjects. (A) Xsens MTx, (B) Regions where the sensors are attached [33].

The data for this study was generated by four women and four men participating in nineteen distinct planned activities for five minutes each. Table I lists the activities carried out by the topics.

Table I. The study involved 19 different activities.

Activity Code	Name of The Activity
A1	the act of sitting
A2	the act of standing
A3	recumbency exercise
A4	Lie down on your right side
A5	transferring to the second floor
A6	descending to the lower level
A7	in the elevator, standing activities
A8	while the elevator is going, standing activities
A9	In the parking lot, there is a lot of walking going on.
A10	Walking exercise on a treadmill at a speed of 4km/h parallel to the ground
A11	Walking on the treadmill at a 15-degree inclination to the ground at a speed of 4 km/h
A12	Running at a speed of 8 km/h is a high-intensity activity.
A13	Running at an 8-kilometer-per-hour pace is a high-intensity activity.
A14	riding on an elliptical machine
A15	Cycling in a horizontal position is an activity that involves cycling in a horizontal position.
A16	Vertical cycling is an activity that involves cycling in a vertical position.
A17	rowing exercise
A18	jumping exercise
A19	basketball game in progress

III. Methodology

A. Person Identification By MobileNet Deep Transfer Learning Technique

The figure below shows the recommended method of identifying people using the MobileNet deep learning method. A brief introduction to the process performed at each stage.

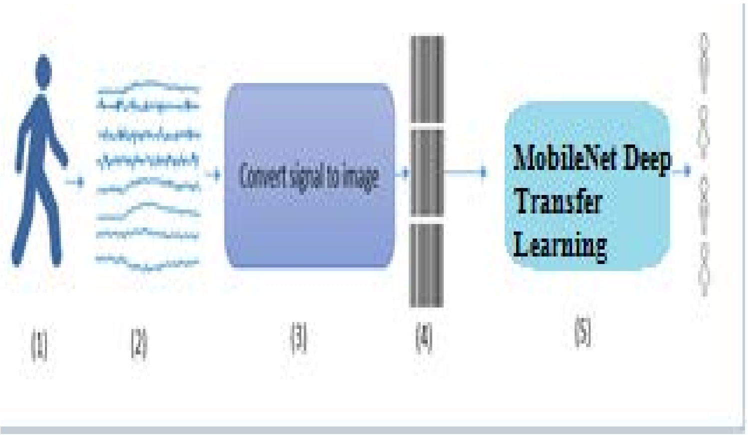

Figure II MobileNet deep transfer learning approach template for person identification (PI-MobileNet)

Block 1: A wearable sensor is attached to the subject at this point. The Xsens MTx sensor devices are placed throughout the subject in five separate locations.

Block 2: A total of 45 signal channels from 5 different places of the body are recorded using accelerometer, gyroscope, and magnetometer sensors.

Block 3: Each action in Table I has a 5-minute signal measurement. These signals are then separated into 5 seconds. The signal length is 5x25 = 125 because the sampling rate is 25. First, use the following equation to convert the value of the signal to a value between 0-255. Since the number of channels is 45, you get an image in the form of 125x45.

$$New\ X_i = round\left(\left[\frac{X_i - Min(X)}{Max(X) - Min(X)}\right] x255\right) \quad (1)$$

Figure III, for example, depicts the image formed by the standing activity signal for each individual.

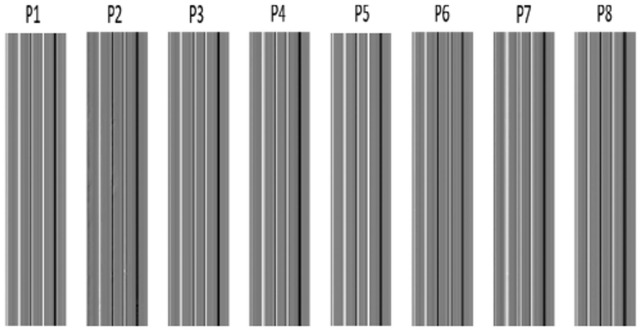

Figure III. For each subject, images of the signals of sitting activity.

Block 4: The image that was made is shown. These photos are sent to the Deep Transfer Learning algorithm for restoration.
Block 5: The image is classified by MobileNet depth transfer learning technology. All architectures of version 1 and 2 of MobileNet have been tested.

B. Deep Trasnfer Learning

AlexNet [25], GoogLeNet [26], VGG Net [27], ResNet [28], and NASNet [29] are just a few of the CNN architectures that have been developed in recent years. All of these architectures have been pre-trained on over one million photos from ImageNet's Large-Scale Visual Recognition Challenge [30]. To transmit these learnings, we can use the model's weights and biases. One of the pre-trained models is the MobileNet architecture [31]. The distinction between MobileNet and other published designs is that MobileNet has a lesser number of parameters. With separable convolution, MobileNet exploits the depth introduced in to lower the computational requirements of the first layers [32]. Latency and precision are trade-offs in MobileNet efficiency. Tasks based on this model can be completed on the CPU without the need of the GPU due to MobileNet's light weight. They, like other large-scale models, can be utilized for classification, detection, and segmentation.

C. The Architecture Of MobileNet

Deep learning networks produce effective outcomes in areas like image analysis and computer vision [5]. There are many hidden layers in a DL. The values given as input to these layers form outputs by going through mathematical operations. These outputs are given as input to the next layer (Figure-1). DL is trained on the designated classes for image analysis and classification. This training requires a previously labeled dataset that DL will work on. Because the analysis of images requires high computational power, configuring the system in which the DL solution will run is very important for performance.

MobileNet is built on a separable depth convolution with two basic layers: depth and point convolution. Deep convolution is the process of filtering the input without adding additional functions. As a result, point-by-point convolution is used to combine the process of creating new features. Finally, a depth separable convolution is the result of combining the two layers. The model applies a single filter to each input channel using deep convolution and then creates a linear combination of deep layer outputs using 1x1 convolutions (point by point). Batch normalisation (BN) and changed long measure (ReLU) liner unit used once every convolution.

Looking at Table II, the MobileNet architecture is seen in detail. The picture given to the model is transformed into the output image by going through the stages in the table. ReLu was chosen as the activation function. This model has approximately 4 million parameters, which is significantly less than the other models.

Table II. The Architecture Of MOBILENET

MobileNet
Input layer
Convolutional layer
Depthwise n Convolution layer Batch Normalization ReLU + Pointwise n Convolution layer Batch Normalization ReLU (n = 1, 2, 3, …, 13 layers)
Global Average Pooling layer
Reshape layer
Dropout layer
Convolutional layer
Softmax layer
Reshape layer
Output

IV. Experiment

In this study, we use a lightweight network called MobileNet to develop a deep neural network architecture that uses a separable deep convolution.

Although the design employs fewer parameters than previous efficiency models at the same level, Google designed MobileNet with the goal of striking a balance between latency and precision and displaying impressive performance. We propose our upgraded MobileNet, which is appropriate for personal recognition categorization utilizing wearable sensor inputs. Starting with the primary convolutional layer, followed by thirteen deep convolutional layers, and at last a degree convolutional layer, our design is galvanized by the MobileNet network layer. once every depth and purpose convolutional layer, use the batch standardization (BN) and changed linear measure (ReLU) trigger functions. the worldwide average grouping layer are accustomed minimize the dimensions of the extracted feature map once all of the convolutional layers have extracted options from the input image.

The Reshape layer, Dropout layer, convolutional layer, Softmax activation perform layer, and thus the Reshape layer, which make up the last 5 layers of the quality MobileNet, are replaced by the Dropout layer and therefore the completely connected layer, which uses the Softmax activation perform. Our fully linked layer will generate more accurate predictions for each class than the preceding five layers using standard MobileNet. Increasing the number of convolutional layers in the model will help it extract more alternatives from the computer file in general.

However, there are some limitations to adding them. We discovered in this work that traditional MobileNet overfits, causing the model to misclassify personel recognition. As a result, our proposed design is capable of resolving this issue. Our modified model's total parameters are reduced, which reduces the time it takes to calculate the model. Table III shows the specifics of our architecture.

Table III. Our MobilNet Articheture

MobileNet
Input layer
Convolutional layer
Depthwise n Convolution layer Batch Normalization ReLU + Pointwise n Convolution layer Batch Normalization ReLU (n = 1, 2, 3, …, 13 layers)
Global Average Pooling layer
Dropout layer
Fully Connected layer
Output

MobileNet is a pre-trained model that recognizes the shape and part of an object in its initial layer and was trained using the ImageNet dataset. So in this paper, we chose all layer deep transfer learning for traditional MobileNet and compared it to our modified MobileNet, which is only two-thirds of the layer ratio and the remaining layers configure the Learn migration. We used our data set and settings to train the model (such as weights, bias, and learning rate). Finally, we assessed MobileNet and our model using the original data set from the previous experiment, without any data augmentation or sampling techniques. MobileNet and our model are tested in the second experiment utilizing data that has been completed by sampling and augmenting data.

V. Results

Signals from 19 different activities of 8 people were used to create the data set for this investigation (4 men and 4 women). Each activity is broken down into 60 segments. As a result, there are 19x8x60 = 9120 signal matrices in the data set. MobileNet deep transfer learning technology is utilized after these signal matrices have been converted into images. 9120 photos were extracted as a consequence of to see if our system worked. There are two MobileNet architectures in operation. The success rate is determined as follows:

$$\mathbf{100} * \frac{\# \, True \; classified}{\# \, True \; classified + \# \, False \; classified} \; (\%) \qquad (2)$$

Table IV shows the success rate of personnel identification using MobileNet Version 1 (MobileNetV1) and MobileNet Vesiyon 2 (MobileNetV2) architectures.

Table IV. Person Recognition Success Of mobilet Networks

Model	**Success rate%**
MobileNetV1	96,82
MobileNetV2	**98,46**

It is seen in Table IV that the person recognition problem with the data obtained from the sensor signals is successfully solved with MobileNet networks. Although success rates are close to each other, MobilnetV2 networks were more successful in person recognition problem than MobileNetv1 networks.

Table V. Success Rate of Mobilenet networks by Physical Activity

Activity	**MobileNetV1**	**MobileNetV2**
A1	100	64,5
A2	100	12,5
A3	97,9	25
A4	93,75	47,916
A5	93,7	18,7
A6	95,8	22,9
A7	97,9	6,2
A8	62,5	16,6
A9	95,8	4,1
A10	100	10,4
A11	100	16,6
A12	100	16,6
A13	95,8	95,8
A14	97,9	22,9
A15	91,6	18,7
A16	95,8	16,6
A17	87,5	27
A18	79,1	27
A19	75	56,2

We also tried our model from the physical activity recognition problem to measure the success of our network after the person recognition process. In general, it has been determined that the MobileNet V1 network is more successful than the physical activation recognition problem. It has shown 100% success in recognizing some physical movements. Detailed results are presented in Table V.

Table VI. Success Rate of Mobilenet networks by region of sensors

Region	MobileNetV1	MobileNetV1
Chest	94,5	93,8
Right Arm	92,1	92,1
Left Arm	91,1	89,9
Right Leg	93,6	92,3
Left Leg	91,9	92

Finally, we have tested the success of our model according to the region where the sensors receive signals. Both models showed high success. In addition, MobileNetv1 networks have a higher success rate. Data from chest-level sensors was more distinctive than data from other regions. Details can be seen in Table VI.

VI. Conclusion

Several biometric technologies have been developed in recent years. Face, voice, fingerprints, palm print, ear shape, and gait are all biometric technologies that have been widely used in security systems. However, because they may be imitated, most of these systems have glaring faults. To address these issues, a new biometric system based on medical signals has been developed. A biometric approach was developed in this study to identify people using wearable sensor inputs. The major goal of this study is to show that signals from portable sensors such as accelerometers, gyroscopes, and magnetometers may be used to identify people. In the future studies, due to the increasing use of mobile devices, person recognition will be done with a deep transfer learning approach through data obtained from portable mobile devices.

References

[1] B. Ngugi, A. Kamis, and M. Tremaine, "Intention to use biometric systems," e-Service Journal: A Journal of Electronic Services in the Public and Private Sectors, vol. 7, no. 3, pp. 20-46, 2011.

[2] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of artificial" gummy" fingers on fingerprint systems," in Optical Security and Counterfeit Deterrence Techniques IV, 2002, vol. 4677: International Society for Optics and Photonics, pp. 275-289.

[3] J. Yu, C. Fang, J. Xu, E.-C. Chang, and Z. Li, "ID repetition in Kad," in 2009 IEEE Ninth International Conference on Peer-to-Peer Computing, 2009: IEEE, pp. 111-120.

[4] G. Gainotti, "Laterality effects in normal subjects' recognition of familiar faces, voices and names. Perceptual and representational components," Neuropsychologia, vol. 51, no. 7, pp. 1151-1160, 2013.

[5] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," IEEE Access, vol. 2, pp. 1530-1552, 2014.

[6] D. Rodrigues, G. F. Silva, J. P. Papa, A. N. Marana, and X.-S. Yang, "EEG-based person identification through binary flower pollination algorithm," Expert Systems with Applications, vol. 62, pp. 81-90, 2016.

[7] S. Marcel and J. d. R. Millán, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation," IEEE transactions on pattern analysis and machine intelligence, vol. 29, no. 4, pp. 743-752, 2007.

[8] Y. Dai, X. Wang, X. Li, and Y. Tan, "Sparse EEG compressive sensing for web-enabled person identification," Measurement, vol. 74, pp. 11-20, 2015.

[9] Y. Sun, F. P.-W. Lo, and B. Lo, "EEG-based user identification system using 1D-convolutional long short-term memory neural networks," Expert Systems with Applications, vol. 125, pp. 259-267, 2019.

[10] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," Pattern recognition, vol. 38, no. 1, pp. 133-142, 2005.

[11] M. Deng, C. Wang, M. Tang, and T. Zheng, "Extracting cardiac dynamics within ECG signal for human identification and cardiovascular diseases classification," Neural Networks, vol. 100, pp. 70-83, 2018.

[12] A. Goshvarpour and A. Goshvarpour, "Human identification using a new matching pursuit-based feature set of ECG," Computer methods and programs in biomedicine, vol. 172, pp. 87-94, 2019.

[13] K. Su et al., "Human identification using finger vein and ECG signals," Neurocomputing, vol. 332, pp. 111-118, 2019.

[14] F. Sufi and I. Khalil, "Faster person identification using compressed ECG in time critical wireless telecardiology applications," Journal of Network and Computer Applications, vol. 34, no. 1, pp. 282-293, 2011.

[15] W. Chang, H. Wang, G. Yan, and C. Liu, "An EEG based familiar and unfamiliar person identification and classification system using feature extraction and directed functional brain network," Expert Systems with Applications, vol. 158, p. 113448, 2020.

[16] R. San-Segundo, R. Cordoba, J. Ferreiros, and L. F. D'Haro-Enriquez, "Frequency features and GMM-UBM approach for gait-based person identification using smartphone inertial signals," Pattern Recognition Letters, vol. 73, pp. 60-67, 2016.

[17] R. San-Segundo, J. D. Echeverry-Correa, C. Salamea-Palacios, S. L. Lutfi, and J. M. Pardo, "I-vector analysis for gait-based person identification using smartphone inertial signals," Pervasive and Mobile Computing, vol. 38, pp. 140-153, 2017.

[18] N. V. Boulgouris, K. N. Plataniotis, and E. Micheli-Tzanakou, Biometrics: theory, methods, and applications. John Wiley & Sons, 2009.

[19] Z. A. A. Alyasseri, A. T. Khader, M. A. Al-Betar, and O. A. Alomari, "Person identification using EEG channel selection with hybrid flower pollination algorithm," Pattern Recognition, vol. 105, p. 107393, 2020.

[20] Ş. Kılıç, Y. Kaya, and I. Askerbeyli, "A New Approach for Human Recognition Through Wearable Sensor Signals," Arabian Journal for Science and Engineering, vol. 46, no. 4, pp. 4175-4189, 2021.

[21] Ş. Kiliç, İ. Askerzade, and Y. Kaya, "Using ResNet Transfer Deep Learning Methods in Person Identification According to Physical Actions," IEEE Access, vol. 8, pp. 220364-220373, 2020.

[22] K. Altun, B. Barshan, and O. Tunçel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," Pattern Recognition, vol. 43, no. 10, pp. 3605-3620, 2010.

[23] B. Barshan and M. C. Yüksek, "Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units," The Computer Journal, vol. 57, no. 11, pp. 1649-1667, 2014.

[24] K. Altun and B. Barshan, "Human activity recognition using inertial/magnetic sensor units," in International workshop on human behavior understanding, 2010: Springer, pp. 38-51.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, vol. 25, pp. 1097-1105, 2012.

[26] C. Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1-9.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[29] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8697-8710.

[30] O. Russakovsky et al., "Imagenet large scale visual recognition challenge," International journal of computer vision, vol. 115, no. 3, pp. 211-252, 2015.

[31] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[32] L. Sifre and S. Mallat, "Rigid-Motion Scattering for Image Classification. arXiv 2014," arXiv preprint arXiv:1403.1687.

[33]O. Dobrucalı , B. Barshan. (2013) Sensor-Activity Relevance in Human Activity Recognition with Wearable Motion Sensors and Mutual Information Criterion. In: Gelenbe E., Lent R. (eds) Information Sciences and Systems 2013.

[34] M. Feurer and F. Hutter, "Hyperparameter optimization," in Automated machine learning: Springer, Cham, 2019, pp. 3-

[35] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," Annual review of biomedical engineering, vol. 19, pp. 221-248, 2017.

[36] D. Hendrycks, K. Lee, and M. Mazeika, "Using pre-training can improve model robustness and uncertainty," in International Conference on Machine Learning, 2019: PMLR, pp. 2712-2721.