



*Correspondence:
Mehshan Ahad,
COMSATS University,
Islamabad, Pakistan,
mehshan.ahad7@gmail.com
com

Pedestrian Gender Recognition with Handcrafted Feature Ensembles

Mehshan Ahad¹, Muhammad Fayyaz²

COMSATS University, Islamabad, Pakistan, ¹mehshan.ahad7@gmail.com,
²fayyazawan@ciitwah.edu.pk

Abstract

Human gender recognition is one the most challenging task in computer vision, especially in pedestrians, due to so much variation in human poses, video acquisition, illumination, occlusion, and human clothes, etc. In this article, we have considered gender recognition which is very important to be considered in video surveillance. To make the system automated to recognize the gender, we have provided a novel technique based on the extraction of features through different methodologies. Our technique consists of 4 steps a) preprocessing, b) feature extraction, c) feature fusion, d) classification. The exciting area is separated in the first step, which is the full body from the images. After that, images are divided into two halves on the ratio of 2:3 to acquire sets of upper body and lower body. In the second step, three handcrafted feature extractors, HOG, Gabor, and granulometry, extract the feature vectors using different score values. These feature vectors are fused to create one strong feature vector on which results are evaluated. Experiments are performed on full-body datasets to make the best configuration of features. The features are extracted through different feature extractors in different numbers to generate their feature vectors. Those features are fused to create a strong feature vector. This feature vector is then utilized for classification. For classification, SVM and KNN classifiers are used. Results are evaluated on five performance measures: Accuracy, Precision, Sensitivity, Specificity, and Area under the curve. The best results that have been acquired are on the upper body, which is 88.7% accuracy and 0.96 AUC. The results are compared with the existing methodologies, and hence it is concluded that the proposed method has significantly achieved higher results.

Keywords: Handcrafted Features, Feature Ensembles Pedestrian Gender Recognition, and Visual Surveillance

1. Introduction

In the current era, technology is growing at a rapid pace, and there can be seen exponential growth in the field of identity recognition in image processing (Ross, L., & Russ, J. C., 2011). Many areas like surveillance (M. Simonelli and A. Quaglio, 2015), Human activities recognition (Saba, T., Rehman, A., Jamail, N. S. M., et al., 2021),

telecommunication, HCI (human-computer interaction) and image processing, etc., are covering most of the technical areas. We can see the economic significance of the field by seeing around the commercial as well as law enforcement applications worldwide. There are so many necessities to overcome the difficulties of making automated systems to process objects in images and extract some useful data in the form of information that leads us to understand information. The new era will be fully automated, which means there would be much involvement in artificial intelligence (Russell, S., & Norvig, P., 2002). Artificial intelligence involves the processes that are done by computer machines to achieve the intelligence of a human. The process involves acquiring information and states the rules and steps to use the information in a particular way so that it may take a wise decision. There are many challenging problems in every field in which AI (Artificial intelligence) and machine learning are necessary to get more accurate results in time effective way (Sharif, M. I., Khan, M. A., Alhussein, M., Aurangzeb, K., & Raza, M., 2021; Amin, J., Sharif, M., Raza, M., Saba, T., Sial, R., & Shad, S. A., 2020; Naz, J., Sharif, M., Yasmin, M., Raza, M., & Khan, M. A., 2020; Khan, M. A., Akram, T., Sharif, M., Javed, K., Raza, M., & Saba, T., 2020; Amin, J., Sharif, M., Gul, N., Raza, M., Anjum, M. A., Nisar, M. W., & Bukhari, S. A. C., 2020). Such areas include medical systems analysis (Khan, M. A., Kadry, S., et al., 2021; Amin, J., Sharif, M., et al., 2020; Sharif, M., Amin, J., et al., 2020) (e.g. brain tumor analysis (Sharif, M., Amin, J., et al., 2020a; Sharif, M., Amin, J., et al., 2020b; Sharif, M., Amin, J., et al., 2020c; Amin, J., Sharif, M., et al., 2020d; Amin, J., Sharif, M., et al., 2019, April; Amin, J., Sharif, M., et al., 2020e) human surveillance (Humans actions (Khan, M. A., Sharif, M., et al., 2020), human gender analysis (Saba, T., Bokhari, S. T. F., et al., 2018; Fayyaz, M., Yasmin, M., Sharif, M., & Raza, M., 2021; Raza, M., et al., 2018). pedestrian attributes (Raza, M., Sharif, M., et al., 2018; Raza, M., Chen, Z., et al., 2018), and person re-identification (Fayyaz, M., Yasmin, M., et al., 2020; Rehman, S. U., Chen, Z., et al., 2018; Rehman, S., Chen, Z., et al., 2016, July; Hussain, S. J., Chen, Z., et al., 2016, January, etc.), biometrics (Sharif, M., Raza, M., et al., 2019; Azeem, A., Sharif, M., et al., 2015), and Satellite imaging (Sharif, M., Khan, S., et al., 2019, April) etc. When it comes to image processing, from the past few decades, there are many complications in the analysis of pedestrians' attributes, such as their behavior analysis, human gender predictions, gait recognition, face recognition, etc. In surveillance, there is a necessity to identify the gender of humans, to do it through traditional way such as manually verify the gender of a person in some crowded area would be a hectic job, and there is much possibility for the occurring of human error, and there could be redundancy of information for a certain person.

Gender recognition has always been a most eye-catching problem for a long time. The initial first technique was the SEXNET (Golomb, B. A., Lawrence, D. T., & Sejnowski, T. J., 1990, November) that uses faces to identify the gender of human beings. This problem later caught the attention of many other researchers, but their all work was constrained most of the human face analysis. There are many scenarios where it is very hard to recognize gender from face images, and this approach does not seem to be fit appropriate there. These scenarios mostly exist when scenes are too obtrusive to capture or take images of the face when there is low-resolution picture is captured, and there might be a possibility of the existence of occlusion there, for example, there could be some glass reflection, the person wearing a mask or facial hair, or some long hair is covering up the face. Some researchers opted for the technique of Gait analysis (Chen, S., Lach, J., Lo, B., & Yang, G. Z., 2016) to recognize human gender, and still, it is an open topic to describe human gait through captured images. It is also very complicated to create such a technique or automated

recognition system to predict gender from the appearances of the human body.

Gender is one of the most important attributes whenever there is a topic about a pedestrian (Raza, M., Zonghai, C., Rehman, S. U., & Shah, J. H., 2017, August) in different applications like multimedia retrieval systems etc. This problem of recognizing the gender of humans has always been a challenging problem that can be observed by seeing the pedestrian images captured at different angles, different environments with the use of surveillance cameras, and poor quality images captured from far distances. The factors that made the recognition very challenging involves the angle variation, i.e., the image captured from the wrong angle or a human pose at a different angle, there could be cluttered backgrounds such as clothes and background are similar, lighting effect could make some images disastrous, and one cannot get enough information out of it. Object occlusion (Enzweiler, M., Eigenstetter, A., et al., 2010, June), like a person covering some other person's body or some car, could get in front of someone, making it hard to recognize a person.

The purpose of gender recognition from pedestrians is to recognize the person as male or female to find the audience in numbers or find out any potential threats before it occurs. There is a need for an extensively big dataset that must acquire various people in images in different environmental conditions and situations. Therefore, we have selected the PETA (Li, D., Zhang, Z., Chen, X., et al., 2016) dataset, which is very diverse, has a huge collection of pedestrian images and has open access. These pedestrian images have several annotated attributes, such as hats, jeans, gender, muffler, shorts, sunglasses, t-shirts, trousers, backpack, age, jacket, long hair, etc. The dataset results will be evaluated in terms of accuracy. Once the image is captured, the human is localized in a rectangle which also represents bounding boxes. Figure 1 demonstrates a few samples of the bounding box from images carried out using annotations provided along with the PETA dataset.

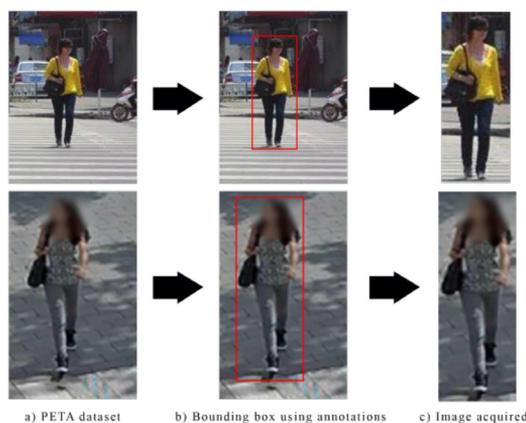


Fig. 1. Various randomly selected pedestrians (Schumann, A., & Stiefelhagen, R., 2017).

The Motivation behind this is to explain a novel technique to recognize the gender of the human being. Gender is a very important attribute of humans. In surveillance, there is always a necessity to detect the gender of a person passing by and gather the information of humans from images in terms of the other attributes and accurately recognize the information which the system is intended to recognize. Due to the evolution of high-quality cameras, there is a vast scope in this field; many systems are

getting automated; for example, a person is usually identified nowadays through facial recognition or fingerprints rather than having traditional data storage. Many offices have automated entrances such as they need to verify their identity by looking at a screen and system identifies the person by doing retina scan. These all biometrics (Almudhahka, N., Nixon, M., & Hare, J., 2016, February), like fingerprints, face, and retina, etc., are used to detect or recognize a human or his particular attribute. Recognition of gender from pedestrians has always been an attractive topic for researchers. It will be very hectic, and errors occurred if some individual has been hired to keep track of pedestrians and recognize them as male or female. To discover the solution to this complication is to maximize the efficiency of security and minimize the security risks in the field of surveillance, a technique is implemented to recognize the gender of humans more precisely and accurately. However, implementing a highly accurate technique has always been complicated due to the various problems involved in it. Gender is the most significant attribute of a human being. To determine the gender of people walking by in daily routine or to keep track of the audience gender-wise is a very demanding problem. There can be observed so much diversity in the appearances and postures variations in the images of pedestrians. Several surveillance systems are deployed in several crowded places like shopping malls, railway stations, airports, parks, etc. surveillance cameras record the captured images with different surroundings and environments for public safety and other security factors. Some are very much exposed to the light or suffered through illumination in some parts of the day due to sun or environmental changes. They can contain low and poor quality whenever they must capture something from a far distance. Through the surveillance, there are huge video data gathered through these systems, and to keep track of people's information through different video analytic tools like pedestrian identification or face recognition etc., have been developed. The major challenges that are faced during gender recognition in pedestrians are due to field of view (FOV). In addition to that, some other issues are listed below.

Occlusion: Features of the human body are not visible due to shadow or another person blocking—for example, a person looking in another direction.

Appearance: Due to dress appearance, gender may appear different if predicted upon clothes.

Illumination (Munir, A., Hussain, A., et al., 2018): Lighting conditions or exposure to more light makes it complicated to recognize the human body, face, or posture.

Therefore, increase the accuracy of multiple human detection and classification of activities is a challenging task. Therefore, increasing the accuracy of gender recognition in pedestrians is a challenging task. Therefore, it is a very meaningful yet challenging topic in the computer vision field.

2. Related work

Gender recognition through facial imaging is the most widely recognized approach. By utilizing face images to deal with perceive gender recognition, Basha et al. (2012) propose a novel way where the nonstop wavelet change is utilized to play out the component choice from every image, and SVM being a direct part, arranged the information as male or female. This approach takes less time contrasted to other order approaches. Also, their approach does fine in the images consisting of varieties in lights and outward appearance, present points, maturing impacts, and so on. Shan et al. (2012) utilize LBP, i.e., local binary patterns, to portray countenances by utilizing the AdaBoost approach to choose different local binary patterns, i.e., LBP features. The obtained accuracy after applying SVM and LBP features is 94.81%. Li et al. (2012)

identify gender (male and female) by using face features including nose, eyes, mouth, temple, foreheads. The issue in their methodology is that multifaceted foundations influence the component removal approach. Considering the various pieces of facial component removal, the gender recognition order through facial could be isolated to the neighborhood include removal and worldwide element removal approaches. The nearby component removal approach concentrates features from evident face focuses like eyes, nose, and mouth, though the global features removal strategy concentrates features from the entire face instead of separating features from facial focuses. Dong et al. (2003) exhibited another methodology utilizing eye-brows to group males and females. They utilized shape-based eyebrows for biometric acknowledgment and recognition of gender. They used three distinct characterization approaches: MD classifier, Least Discriminant Analysis (LDA) (Gromski, P. S., Muhamadali, H., et al., 2015), and SVM (Meyer, D., & Wien, F. T., 2015). These strategies were strained on images from two different openly accessible databases of the facial image. Amayeh et al. (2018, March) proposed that the shape of the hand is a noticeable component for male and female recognition grouping and portioned this feature outline into 6 distinct sections compared to the hand's palm and the fingers. To see each section, contour, area, and limit features were used dependent on Fourier descriptors as well as Zernike minutes. Likewise, they also figured the separation of a given section from two distinct Eigen spaces for order; both belong to the gender, i.e., from the male and female classes. 98% was the distinguishing identification rate of the methodology. Cho et al. (1977) utilize EMD to recognize various individuals dependent on impression. The strategy is achievable when clients have covered hand means wear the glove on the hand. Cao et al. (1998) fabricate the arrangement of perceiving male and female from full self-perceptions; his research was the principal endeavor to gender acknowledgment utilizing static human self-perception. By incorporating the section-based portrayal and troupe learning calculations, PBGR (Cao, C., Schultz, A. B., et al., 1998), i.e., Part-based gender recognition approach, is proposed to arrange male and female recognition utilizing a solitary front-view or a back-view image with 75.0% accuracy. Gender Recognition dependent on static body features can perform distinguishing identification. Be that as it may, since individuals always show signs of change in the outlook, styles, and areas, some proposed using conduct features for male and female classification, such as the development of the action. Yu et al. (2014) right off the bat show a test wherein they solicited members to recognize the gender recognition from moving human outlines. At that point, their Appearance Based male and female, i.e., gender arrangement, was upgraded through the removed humanoid learning from the analysis. In their research, they take the CASIA Database. The consequence of the research showed that walk features could be very helpful to improve the accuracy of the male and female recognition classification. Moreover, Abdenour et al. (2009) looked at different plans for facial acknowledgment. Their exploratory outcomes demonstrated that the blend of movement and appearance was helpful for gender recognition for the well-known countenances. Their examination evaluated the efficient presentation of the LBP for depicting and investigating appearances depends on three different video databases. 90.3%, 78.3%, and 88.7%, accuracies were attained individually by utilizing the VLBP-based approach. Williams et al. (2009) new mechanized male and female recognition while occupied with moving action. The strategy with pre-processing phase utilizing rule part investigation and AdaBoost classifier (Wen, X., Shao, L., et al., 2015) accomplishes a precision of 87.8%. Nonverbal conduct is a significant piece of human associations. The examination (Won, A. S., Yu, L., et al., 2012) already portrayed another strategy for deciding gender

orientation personality utilizing AI with motions taken from Microsoft Kinect (Zhang, Z., 2012). The strategy accomplished 83% accuracy in foreseeing the recognition of male and female, even from exceptionally fewer area exposures, for example, 10 seconds introduction of the members. The gender recognition likewise requires a higher computational multifaceted nature since conduct features need image sequencing for account developments. Both genders, either male or female, have unmistakable inclinations in regards to dressing (Raza, M., Zonghai, C., et al., 2017). In this manner, as indicated by a person's dressing features, Ueki, K. Komatsu, H., Imaizumi, S. et al. (2004, August) proposed a male and female order strategy that incorporates the data from so many pieces of solitary images. By coordinating likely haircut and attire and applying the classifiers like 'PCA' and 'GMM' (Seo, C., Lee, K. Y., & Lee, J., 2001) on a great number of Test images, exploratory outcomes exhibited that the combination procedure fundamentally brought down the rate by 25% in male and female order as compared with the customary methodologies that utilization features of the face as it were. The examination led by Yuan et al. (2016) out of the blue showed the adequacy of footwear appearance for male and female recognition as the component. As per primer test output, they presumed that HOG, i.e., histogram of oriented gradient, which spoke to a footwear image in addition to nonlinear SVM, gives acceptable outcomes with 85.49% accuracy. The Attire features could fill in as and-other kind of features to arrange male and female, which is simpler to get and discriminative even inside low-quality of image. The male and female recognition classifier with a single component accomplishes a grouping precision long from impeccable. The multi-factors imply male and female recognition utilizing a few features collectively. Hadid et al. (2008) consolidated movement and appearance for male and female recognition examination, and they acquired 90.3%, 78.3%, and 88.7% accuracy, respectively, which are testing on three uniquely designed databases. Ueki, K. Komatsu, H., Imaizumi, S. et al. (2004, August) exhibited a male and female recognition strategy that incorporated facial data, haircut, and attire data. They had the option to lessen false groupings make through the traditional methodology that lone utilize face features by 25.1%. Xia et al. (2013, September) explored the blend of surface modalities and the shape for male and female recognition. They played out the investigation in 2 different ways: one breaker the outcomes from dark images, and different wires the outcomes from dim images and 3D networks.

In preprocessing, datasets of the images cannot be utilized directly for further processing when the quality of the image is enhanced (Shah, G. A., Khan, A., et al., 2015). Generally, images of the datasets have skew, noise (Sharif, M., Irum, I., Yasmin, M., & Raza, M., 2017; Irum, I., Shahid, M. A., Sharif, M., & Raza, M., 2015), and other defects that can damage the image shape or varied in size. Many image enhancement algorithms (Agaian, S. S., Silver, B., & Panetta, K. A., 2007; Ansari, G. J., Shah, J. H., Sharif, M., & ur Rehman, S., 2020) are implemented step by step on the images to enhance quality in preprocessing phase to be suitable for further feature extraction and classification. In the preprocessor module, many algorithms for extracting features (Xu, J., Liu, J., Yin, J., & Sun, C., 2016) work with the sequences of the image silhouettes (Dibra, E., Jain, H., Oztireli, C., Ziegler, R., & Gross, M., 2017) instead of raw video sequences. PETA dataset also consists of the images taken from the surveillance camera, and these images are the raw video sequences from which the images are extracted. The extracted images also have annotations along with them so that a bounding box could be made around a person's body and the only interesting object that gets extracted is the human body in our case. Various techniques could be implemented to attain maximum information from the image which it contains, such as

segmentation (Shi, J., & Malik, J., 2000), contrast normalization (Buhr, J. D., Goodwin, R. M., et al., 2000), background subtraction (Piccardi, M., 2004, October), denoising (Buades, A., Coll, B., & Morel, J. M., 2005, June), and morphology (Dougherty, E., 2018), etc.

Features can be defined as certain positions in images, including highland's peak, turning angles of buildings, and many other distinctive characteristics according to the application (Saba, T., et al., 2019; Khan, M. A., et al., 2019; Rauf, H. T., Saleem, B. A., et al., 2019). The certain features mentioned above are famous as Key-point Features. The detection of these features is known as key-point detection, used to sense the key-point features and produce a Feature Descriptor. A descriptor defines the image area around the desired point. For example, Granulometry Based Descriptors (Ferrari, S., Piuri, V., & Scotti, F., 2008, July), Oriented Fast and Rotated Brief and HOG, etc. In this work, we have focused on HOG, Gabor filter (Liu, C., & Wechsler, H., 2002), and Granulometry (Vincent, L., 1994) in this study. The granulometry method can be described as the process where different opening results detect any object's characteristics, such as shape and size. The technique can be applied to dual images and suitable for distinguishing the objects easily recognizable. Esteban et al. (2003) carried out research based on Granulometry based descriptors to distinguish between vigorous regions and pathological areas. The experiments were conducted by examining the granularity of the prostatic flesh and showed an extraordinary result by the descriptor. Bacher, Uwe & Mayer also studied the Granulometry-based descriptor, Helmut (2012), to ascertain the performance of Granulometry features in the process of automated abstraction of roads. The chosen feature was used with small inaugurals to distinguish the Iranian road networks and deliver enough evidence about the objects in high-quality images.

HOG descriptors presented by Dalal and Triggs (2005, June) refer to a bunch of histograms constituting of Pixel alignment provided by certain gradients. The descriptors can increase the radiance variations and typically use the corners of the element. Alterman, R., Zito-Wolf, R., & Carpenter, T. (1998) studied the pragmatic feature descriptor Histogram of Oriented Gradients, also known as HOG, and worked on its suitability with Traffic signs. The researchers raised the dataset of the rapidity limit and carried a standard test for the selected feature. The results showed that HOG features are appropriate for recognizing the traffic signals because of their fast recognition ability and can be implemented without any complication. The feature descriptor was also used in detecting the visuals of rough automobile locations by Papanikolopoulos, N. P., Krantz, D. G., et al. (2003). In their study, the scholars studied the orientation of cars by using HOG and gathered positive results 88% of the time. The study contains each classifier being tracked to identify various vehicles, including the classifiers limited to track angles and turns. Gabor, D. (2007) introduced the Gabor filter in 1946 as the statistically independent Feature descriptor to detect the surface, edges, and elements. Gabor filter was used to recognize the handwritten digits by Hamamoto, Y., Uchimura, S., et al. (1998) in their study and examined the Feature descriptor by carrying experiments with minimum fault rate. Yang, J., Liu, L., Jiang, T., & Fan, Y. (2003) researched to improve the Fingerprint images and compared the Traditional and Modified Gabor Filter. The study consists of experimental outcomes which suggest that the selected Gabor-Filter can decrease the FRR by 2%, showing extraordinary results.

Next, after extracting the feature, the system is classified utilizing appropriate classifiers, and a verification process is implemented based on these selected features. In the human detection process, consider the object as a human or non-

Table 1: Various Feature extraction technique used for pedestrian gender recognition

Reference	Methodology	Results	Dataset/Variety
Cao, L., Dikmen, M., Fu, Y., & Huang, T. S. (2008, October)	HOG	75% Accuracy	View (Frontal, back)
Ge, W., Collins, R. T., & Ruback, B. (2009, December)	PiHOG, color	76% Accuracy	View (Frontal)
Guo, G., Dyer, C. R., Fu, Y., & Huang, T. S. (2009, September)	BIF+PCA/LSDA	80.6% Accuracy	View (Frontal, Back)
Ng, C. B., Tay, Y. H., & Goi, B. M. (2012)	HOG, color histogram, skin pixels	82.4% Accuracy	unconstrained
Ng, C. B., Tay, Y. H., & Goi, B. M. (2012)	HOG-assisted deep feature learning	0.95 AUC	View (mix)

human and in the phase of activities classification by applying the selected classifiers and considered in different classes where class M denotes males, F denotes females. Most popular choices of classifiers are KNN (Yu, Z., Chen, H., Liu, J., et al., 2015), linear SVM (Keerthi, S. S., Shevade, S. K., et al., 2001), Random forest (Liaw, A., & Wiener, M., 2002), AdaBoost (Mathanker, S. K., Weckler, P. R., et al., 2011) and ANN classifier (MageshKumar, C., Thiyagarajan, R., et al., 2011, March).

3. Proposed work for pedestrian gender recognition

In this section, A new system is based on the fusion of extracted features (Ayyaz, M. N., Javed, I., & Mahmood, W., 2016). The proposed algorithm has the following steps: a) Preprocessing, b) image acquisition, c) Feature extraction, d) Feature fusion, and e) Classification. In figure 2, the proposed methodology is illustrated.

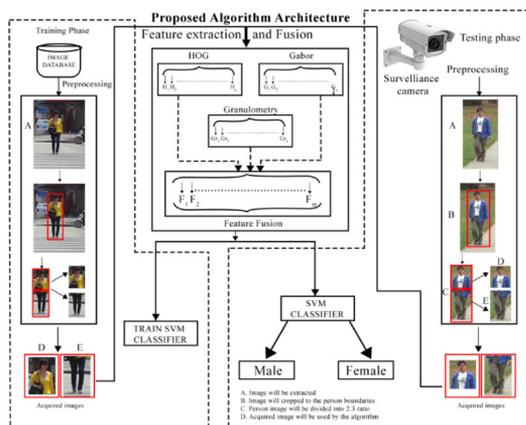


Fig. 2. The Architecture of the proposed methodology

3.1. Preprocessing

Preprocessing always plays a vital role in image processing to fetch out the important details from images. The extracted information, when it comes with more details, becomes helpful in predicting and evaluating results. In this study, to extract maximum information from the dataset's images, we have combined all the datasets such as 3DPeS, CAVIAR4REID, CUHK, GRID, MIT, PRID, SARC3D, Town Center, and VIPeR. The total numbers of images were 19000. Images having bad resolutions and occluded images are removed from the dataset, and the same picture of the person has been removed to remove redundancy that could affect the results.

To obtain the maximum information from the images, we use the annotated dataset; the interesting areas are selected only, providing maximum information regarding that pedestrian in the image, shown in figure 3.

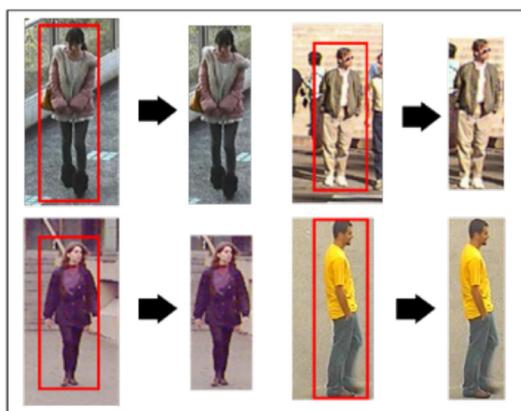


Fig. 3. Image acquisition from bounding boxes after using annotations on PETA dataset

Once the annotated images are fetched out, we have classified the dataset based on their gender attributes, such as males and females. The whole dataset has been divided into categories, the male attribute category has images of 10,137 images, and the female attribute category has 8472 images. Now, there are two possibilities through which we can get more information based on attributes such as accessories, hair, and bag packs, etc. the possibilities are as follows: i) To use all the images as it is in the dataset on the algorithm to evaluate results, ii) crop the images into 60% and 40% as shown in Figure 4.

To crop the image in upper body parts and lower body parts, we have selected the ideal ratio of any human body that is 60% and 40%. In this article, we have separately cropped the male and female images and named those images as upper body male, upper body female and lower body male, and lower body female. The whole process of preprocessing for our work is illustrated in figure 5.

3.2. Feature Extraction

In this work, we have used three feature extractors: HOG, Gabor, and Granulometry, to get the separate feature vectors. We have used different score values for each feature to get feature vectors. We have utilized several score values for each feature separately to get maximum features out of the images. These features are extracted

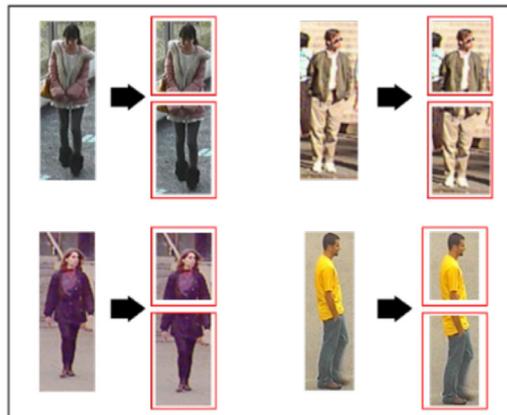


Fig. 4. Cropping image into 60% and 40%

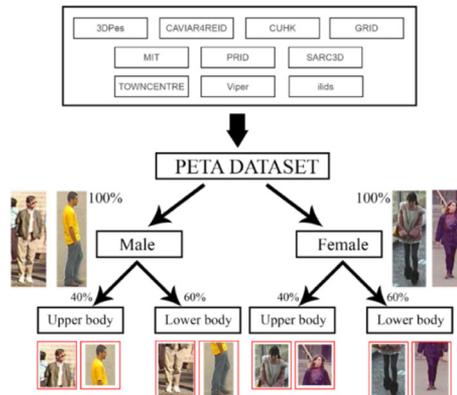


Fig. 5. Representation of preprocessing, which is processed on the dataset

on the PETA dataset variants, which are attained after preprocessing. Once the feature vectors are acquired, they are fused to create a strong feature vector, But it is impossible to acquire such strong features using only one feature extractor. Once the final fused feature vector is obtained, we use SVM and KNN classifiers for training and testing purposes. The whole architecture of getting features extracted and their fusion is illustrated in figure 6.

3.2.1. HOG

One of the well-known features to recognize objects is the Histogram of Oriented Gradients (HOG). In local regions, the intensities of edges are taken through orientation histograms through which HOG features are calculated. HOG applies to image regions from all locations existing on a dense grid, and then these features are classified through classifiers such as Support vector machines SVM. HOG feature descriptor is globally implemented on various domains to characterize and classify shapes through

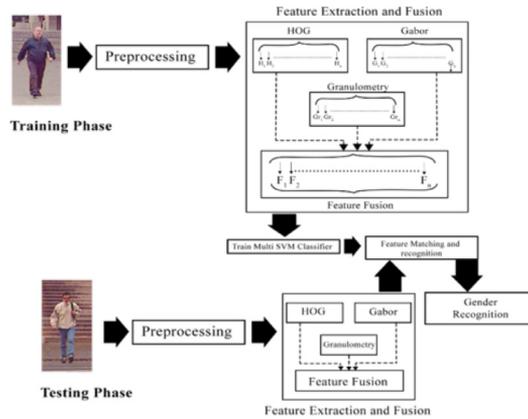


Fig. 6. Architecture diagram of the feature extraction process

different shapes. Through the distribution of local intensities edge directions, a local object's appearances and visual appearance can be determined. HOG is one of the widely utilized feature descriptors for object detection (Rashid, M., Khan, M. A., et al., 2019). The most common example is the human body or face. This is the reason HOG is used in this work to extract the features. The maximum number of features that HOG can acquire is 3780, but in our technique, we have only used 200 features from the image of the dataset. The process of obtaining a feature vector from HOG is illustrated in figure 7.

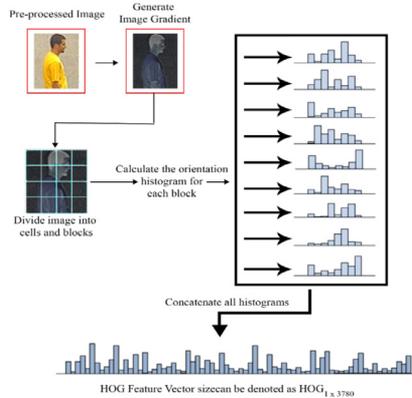


Fig. 7. HOG feature extraction

3.2.2. Gabor filter

Gabor filter responds as a complex set of multi-resolution values when they deal with the raw features. However, to recognize and detect the complex real-world structures in images, these features extracted using Gabor filters can be used. The extraction of the Gabor feature through an image is illustrated below. The maximum

number of features acquired by Gabor is 96, but only 32 features are selected for the selected image in the fusion process. Gabor filters having multiple orientations and scaling are used to form up a filter bank. After putting results of convolution on the input image, results were carried out, and each filter together forms the Gabor features. The illustration of acquiring the feature vector of the Gabor filter is shown in figure 8.

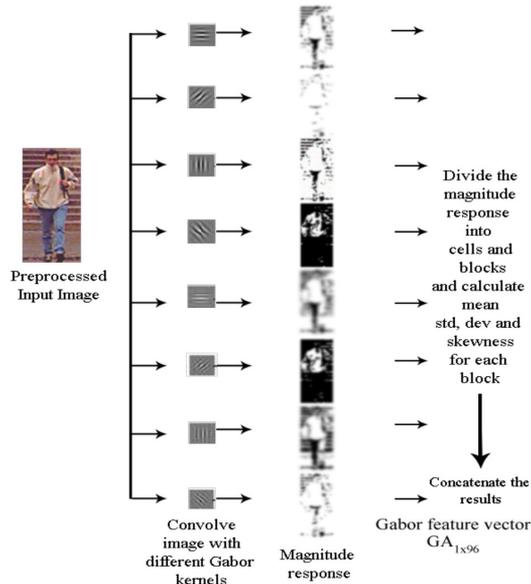


Fig. 8. Gabor Feature extraction

3.2.3. Granulometry Features

Granulometry presented by Prodanov, D., Heeroma, J., & Marani, E. (2006) is the most significant technique used for measuring the magnitude of inorganic scraps in sedimentary rocks and relating their material. This method is used universally by many researchers. The granulometry method can be performed through numerous techniques based on their level of simplicity and advancement. The more straightforward techniques involve definitive sediment filtering (R. L. Tucker, 2000), and the complicated granulometry methods include grain magnitude dimension and laser rays. The total number of feature that granulometry acquire is 78, and in this work, 30 feature score for this feature vector are fused with other feature extractors. The feature vector obtained from granulometry is shown in figure 9 when plotted on an area graph.

3.3. Feature Fusion

In many classification applications, the most important problem is achieving desirable performance if a set of feature vectors are given. Feature fusion (Khan, M. A., Sharif, M., et al., 2020; Saba, T., et al., 2019; Amin, J., Sharif, M., Raza, M., & Yasmin, M., 2018; Arshad, H., Khan, M. A., et al., 2019) is a set of different feature vectors in such an effective way that the resultant vector gives better results. Generally, there are two feature fusion methods: in serial fusion, different feature vectors are merged to create a combined vector, while parallel fusion is performed in parallel format. The equation for obtaining the feature vector of HOG can be denoted as

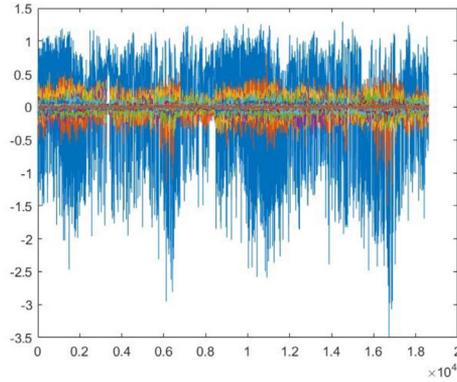


Fig. 9. Area graph of the granulometry feature vector

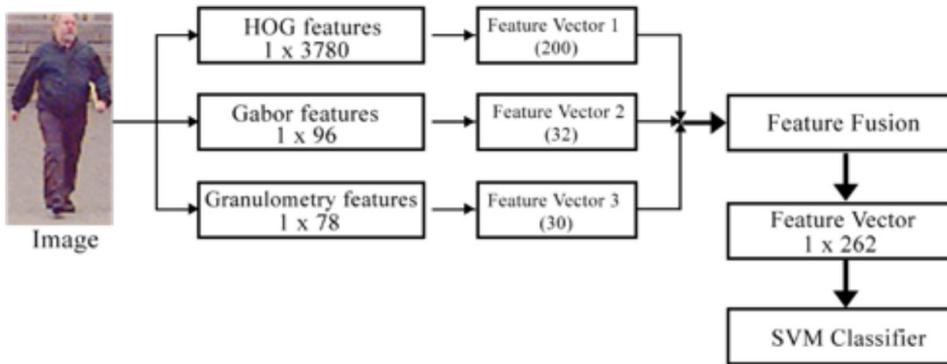


Fig. 10. Flow diagram of feature fusion

$$\text{HOG}_{H_{1 \times 3780}} = \{H_{1 \times 1}, H_{1 \times 2}, H_{1 \times 3} \dots \dots H_{1 \times 3780}\} \quad (3.1)$$

where p is the number of features that HOG will acquire.

Furthermore, the Gabor (3.2) and granulometry (3.3) feature can be denoted as

$$\text{GA}_{g_{1 \times 96}} = \{G_{1 \times 1}, G_{1 \times 2}, G_{1 \times 3} \dots \dots G_{1 \times 96}\} \quad (3.2)$$

$$\text{GR}_{gr_{1 \times 78}} = \{Gr_{1 \times 1}, Gr_{1 \times 2}, Gr_{1 \times 3} \dots \dots Gr_{1 \times 78}\} \quad (3.3)$$

where q and r denote the number of features that Gabor and granulometry will contain, respectively.

The final feature vector will be represented after applying score, i.e., HOG, Gabor, and Granulometry 200, 32, and 30, respectively, and n is the number of images that can be represented as

$$\Psi_{i \times 262} = \sum_{i=0}^n (hog_{i \times 200} + Ga_{i \times 32} + Gr_{i \times 30}) \quad (3.3)$$

The process of acquiring feature vectors from one image after fusing all obtained feature vectors is shown in figure 10.

After calculating the final feature vector, the feature vector will be fed to classifiers to acquire results based on different performance measures.

4. Results and Analysis

In this chapter, the proposed methodology results are gathered using datasets to detect genders from pedestrians. The experiments have been conducted on the PETA dataset. The details of the dataset will be shown below sections. These images are trained and tested on an SVM classifier using 10-folds cross-validation. To evaluate the performance of our algorithm, we have used performance measures which are Accuracy (Acc), Precision (Pre), Sensitivity (Sen), Specificity (Spe), and Area under the curve (AUC). The experiments are evaluated on MATLAB 2018b, using a 64-bit Operating System, 16GB RAM, and a 2.20 GHz processor.

4.1. Results

To evaluate our proposed technique, we have used five performance measures which are illustrated in Table 2.

Table 2: Performance measures for result evaluation

Sr. No.	Performance Measure	Equation
1	Accuracy	$\frac{TP + TN}{FP + TP + FN + TN}$
2	Precision	$\frac{TP}{TP + FP}$
3	Sensitivity	$\frac{TP}{TP + FN}$
4	Specificity	$\frac{TN}{TN + FP}$
5	Area Under Curve	$\int_{-\infty}^{\infty} TPR(T)FPR/(T)dT$

To calculate the performance of the proposed methodology, we have divided our dataset into two classes: male and female. These classes have also been subdivided into 4 classes, i.e., male upper body (MUB), female upper body (FUB), male lower body (MLB), and female lower body (FLB), which contains 10137 images of male and 8827 images of female as shown in Table 3.

Table 3: Datasets used for result evaluation

Dataset	Male	Female
Full Body	10137	8827
Upper Body	10137	8827
Lower Body	10137	8827

The results have been compared using two classifiers KNN and SVM, which show that SVM is better concerning other classifiers. To classify our dataset, we have performed the experiments on datasets separately, and we have assigned a label to the classes of the dataset, which is male and female, as described in Table 4.

Table 4: Classes for classification using classifiers

Classes	Labels

Male full body	MFB
Male upper body	MUB
Male lower body	MLB
Female full body	FFB
Female upper body	FUB
Female lower body	FLB

4.1.1. Experiments

To find the best combinations of features gathered from feature extractors in our case are HOG, Gabor, and Granulometry, we have to conduct experiments and find out the best combination of features from all three feature extractors, which will provide the best accuracies in results. 5 tests are conducted by setting up different feature score values in HOG, Gabor, and Granulometry, as shown in Table 5.

Table 5: Test/Experiments statistics of features score

Test number	HOG	Gabor	Granulometry	Total features
1	100	20	25	145
2	130	28	25	183
3	150	31	28	209
4	200	32	30	262
5	400	33	33	466

These features will apply to all datasets listed in the Table, 30% of the image of each class will be used for training, and 70% will be used for testing purposes. We will use the SVM and KNN classifier for the result classification.

4.1.2. Test 1

Test 1 is conducted on score of HOG = 100, Gabor = 20, and Granulometry = 25 as shown in Table 6.

Table 6: Classification results using test 1 on full-body PETA dataset

Classifier	Evaluation metrics				
	Acc. (%)	Prec. (%)	Sens. (%)	Spec. (%)	AUC (%)
Linear SVM	75.2	75.0	74.5	74.5	82
Quadratic SVM	83.4	83.5	83.0	83.0	91
Cubic SVM	83.6	83.5	83.0	83.0	91
Fine Gaussian SVM	54.1	36.0	50.0	50.0	68
Medium Gaussian SVM	83.9	84.5	83.5	83.5	92
Coarse Gaussian SVM	75.6	76.0	75.0	75.0	83
Fine KNN	79.0	79.0	78.5	78.5	79

Medium KNN	81.1	81.0	81.0	81.0	90
Coarse KNN	78.2	81.0	77.0	77.0	88
Cosine KNN	82.1	82.0	81.5	81.5	91
Cubic KNN	80.5	81.0	80.0	80.0	89
Weighted KNN	81.1	82.0	80.0	80.0	90

The result has been classified through SVM and KNN classifiers, and the performance measures were taken according to the Table, and the results are shown in Table 6.

4.1.3. Test 2

For Test 2, the features extractors score are set upon to 130, 28, and 25 for the feature extractor HOG, Gabor, and Granulometry, respectively. The results after the classification through SVM and KNN are shown in Table 7. The best results are achieved using cubic SVM.

Table 7: Classification results using test 2 on full-body PETA dataset

Classifier	Evaluation metrics				
	Acc.(%)	Prec. (%)	Sens. (%)	Spec. (%)	AUC (%)
Linear SVM	76.3	76.0	75.0	76.0	83
Quadratic SVM	84.1	84.5	83.5	83.5	92
Cubic SVM	84.4	84.5	83.0	84.0	92
Fine Gaussian SVM	54.0	32.0	50.0	50.0	67
Medium Gaussian SVM	84.0	84.0	83.5	83.5	92
Coarse Gaussian SVM	76.6	77.0	76.0	76.0	84
Fine KNN	79.3	79.5	78.5	78.5	79
Medium KNN	81.3	82.0	80.5	80.5	90
Coarse KNN	76.2	81.5	75.0	75.0	89
Cosine KNN	82.5	82.5	82.0	82.0	91
Cubic KNN	80.7	81.0	80.0	80.0	89
Weighted KNN	81.1	82.5	80.0	80.0	90

4.1.4. Test 3

Test 3 is conducted HOG, Gabor, and Granulometry with 150, 31, and 28, respectively. Performance measure results acquired after the classification through SVM and KNN. On cubic SVM, the achieved accuracy is 84.5, and the area under the curve is 0.92. The results are illustrated in Table 8.

Table 8: Classification results using test 3 on full-body PETA dataset

Classifier	Evaluation metrics				
	Acc (%)	Prec. (%)	Sens. (%)	Spec. (%)	AUC (%)
Linear SVM	76.5	76.5	76.0	76.0	84
Quadratic SVM	84.2	84.5	84.0	84.0	92
Cubic SVM	84.5	84.5	84.0	84.0	92
Fine Gaussian SVM	54.0	32.0	50.0	50.0	64
Medium Gaussian SVM	84.0	84.5	82.5	82.5	92
Coarse Gaussian SVM	76.6	77.0	76.0	76.0	84
Fine KNN	79.0	79.0	78.5	78.5	79
Medium KNN	81.0	82.0	80.0	80.0	90
Coarse KNN	75.7	82.0	73.5	73.5	89
Cosine KNN	82.7	83.0	82.0	82.0	91
Cubic KNN	80.0	81.0	79.0	79.0	87
Weighted KNN	80.7	82.5	80.0	80.0	89

4.1.5. Test 4

Feature extractors score values are HOG = 200, Gabor = 32, and Granulometry = 30.

Table 9: Classification results of test 4 on full-body PETA dataset

Classifier	Evaluation metrics				
	Acc. (%)	Prec. (%)	Sens. (%)	Spec. (%)	AUC (%)
Linear SVM	77.3	77.5	77.0	77.0	85
Quadratic SVM	84.1	84.0	84.0	84.0	92
Cubic SVM	85.2	85.5	85.0	85.0	92
Fine Gaussian SVM	54.0	30.0	50.0	50.0	60
Medium Gaussian SVM	84.1	84.5	83.5	83.5	92
Coarse Gaussian SVM	77.3	77.5	77.0	77.0	85
Fine KNN	79.1	79.5	78.5	78.5	78
Medium KNN	80.3	81.5	79.5	79.5	89
Coarse KNN	70.7	81.5	68.0	68.0	89
Cosine KNN	83.1	83.0	83.0	83.0	92

Cubic KNN	79.1	81.0	78.0	78.0	88
Weighted KNN	79.6	82.5	75.0	75.0	89

The result has been classified through SVM and KNN classifiers which are shown in Table 9. The best accuracy of 85.2% and 0.92 AUC is achieved with cubic SVM.

4.1.6. Test 5

Test 5 is shown in Table 10, which is conducted on the score values of HOG = 400, Gabor = 35, and Granulometry = 35.

Table 10: Classification Results of Test 5 on Full body PETA dataset

Classifier	Evaluation metrics				
	Acc.(%)	Prec. (%)	Sens. (%)	Spec. (%)	AUC (%)
Linear SVM	78.5	78.5	78.0	78.0	86
Quadratic SVM	82.7	83.0	82.5	82.5	90
Cubic SVM	84.8	85.0	84.5	84.5	92
Fine Gaussian SVM	54.0	30.0	50.0	50.0	56
Medium Gaussian SVM	84.1	84.5	83.5	83.5	92
Coarse Gaussian SVM	78.3	79.0	78.5	78.5	86
Fine KNN	77.4	78.5	76.5	76.5	76
Medium KNN	75.9	81.0	74.0	74.0	86
Coarse KNN	58.9	78.0	55.0	55.0	85
Cosine KNN	83.5	78.0	83.5	83.5	92
Cubic KNN	79.1	81.0	78.0	78.0	88
Weighted KNN	75.3	81.5	73.5	73.5	86

4.1.7. Best Test Configuration

We have conducted 5 different tests based on the features score for each feature descriptor, i.e., HOG, Gabor, and Granulometry, and fused the best features to form one feature. The test results indicate that the best feature configuration acquired is when HOG=200, Gabor=32, and Granulometry=30.

The best performance measures are achieved through SVM classifiers, i.e., Cubic SVM, Quadratic SVM, and Medium Gaussian SVM, as shown in Table 11 according to the accuracy.

Table 11: Best Classification Results on Tests at Full body PETA dataset

Best SVM Results	Test 1 (Acc.)	Test 2 (Acc.)	Test 3 (Acc.)	Test 4 (Acc.)	Test 5 (Acc.)
Quadratic	83.4%	84.1%	84.2%	84.1%	82.7%

Cubic	83.6%	84.4%	84.5%	85.2%	84.8%
Medium Gaussian	83.9%	84.0%	84.0%	84.1%	84.1%

Hence, the tests have shown that the best configurations for the features are set in test 4, and we will use the same configuration for the rest of the results, which will be taken on the cropped PETA dataset. Figure 11 shows the time results graph taken on the SVM classifier.

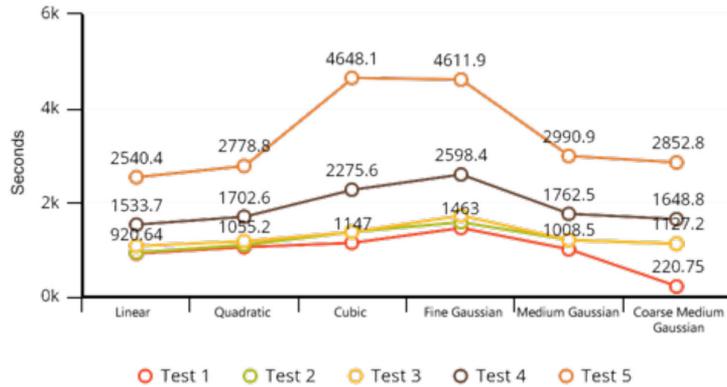


Fig. 11. SVM Classifier Results Time Graph

Results are also evaluated on the KNN classifier; Figure 12 shows the time graph for the KNN classifier takes on the full-body PETA dataset.

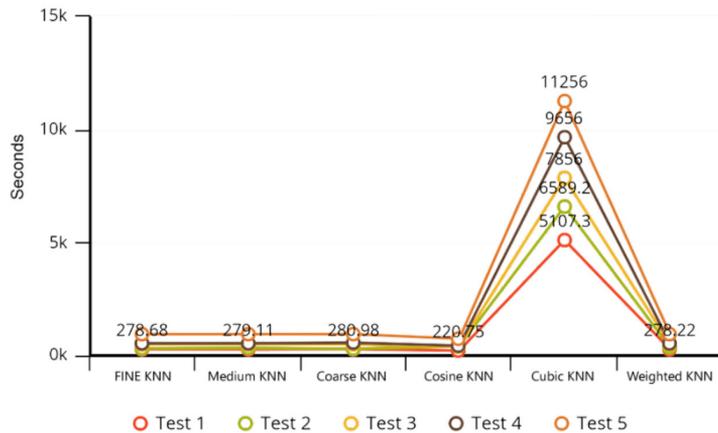


Fig. 12. KNN Classifier Results Time Graph

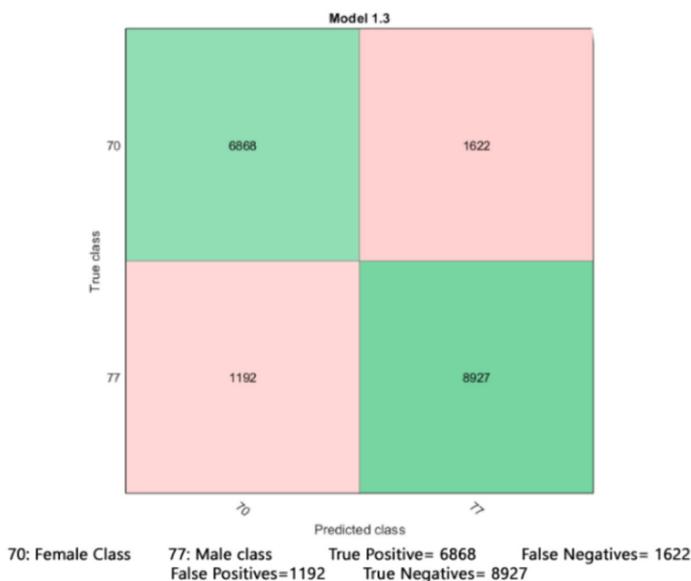


Fig. 13. Confusion matrix of the best result acquired on full-body Dataset

4.1.8. Configuration on Lower Body Dataset

As far, it has been concluded that the best set of features has been acquired through the SVM, as shown in Table 12. The features will remain the same, and we will test it on the lower body dataset, which is 60% of the total image acquired from the PETA dataset.

The experiment results obtained from the lower body dataset reveal that the proposed algorithm improves the accuracy rate compared with other existing methods using HOG, Gabor, and Granulometry features, as shown in Table 12.

Table 12: Classification Results on lower body PETA dataset

Classifier	Evaluation metrics				
	Acc. (%)	Prec. (%)	Sens. (%)	Spec. (%)	AUC (%)
Linear SVM	73.6	73.0	72.5	72.5	80
Quadratic SVM	79.4	79.5	78.5	78.5	87
Cubic SVM	80.0	80.0	79.0	79.0	92
Fine Gaussian SVM	56.9	50.0	50.0	50.0	55
Medium Gaussian SVM	79.6	79.5	78.5	78.5	87
Coarse Gaussian SVM	73.3	73.5	71.0	71.0	80
Fine KNN	73.5	74.0	71.5	71.5	71

Medium KNN	71.0	77.0	62.0	62.0	79
Coarse KNN	57.6	78.5	51.0	55.0	79
Cosine KNN	79.1	78.0	83.5	83.5	87
Cubic KNN	79.1	81.0	78.0	78.0	88
Weighted KNN	70.7	78.5	66.5	66.5	81

The confusion matrix of the SVM Classifier on the lower body PETA dataset is presented in Table 12. It contains two classes that show the percentage of correctly and wrongly classified examples. The confusion matrix and the ROC curve is also shown in figure 13 and 14, respectively.



Fig. 14. Confusion Matrix on lower body PETA dataset

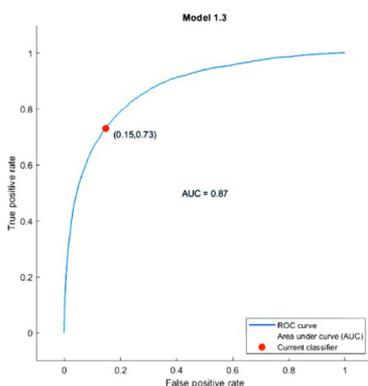


Fig. 15. ROC Curve on lower body PETA dataset

4.1.9. Configuration on Upper Body Dataset

As far, it has been concluded that the best set of features has been acquired through the SVM, as shown in the Table. The features will remain the same, and we will test it on the Upper Body dataset, which is 60% of the total image acquired from the PETA dataset.

The experiment results obtained from the upper body dataset reveal that the proposed algorithm improves the accuracy rate compared with other existing methods using HOG, Gabor, and Granulometry features, as shown in Table 13.

The confusion matrix of the SVM Classifier on the upper body PETA dataset is presented in figure 15. It contains two classes that show the percentage of correctly and wrongly classified examples. The ROC curve for Cubic SVM is also shown in figure 16.

Table 13: Classification Results on Upper body PETA dataset

Classifier	Evaluation metrics				
	Acc. (%)	Prec. (%)	Sens. (%)	Spec. (%)	AUC (%)
Linear SVM	80.7	80.5	80.0	80.0	88
Quadratic SVM	87.4	87.0	87.0	87.0	95
Cubic SVM	88.7	88.5	88.5	88.5	96
Fine Gaussian SVM	54.5	52.5	50.0	50.0	66
Medium Gaussian SVM	87.6	87.5	87.5	87.5	95
Coarse Gaussian SVM	80.6	81.0	80.5	80.5	88
Fine KNN	81.4	82.0	80.5	80.5	81
Medium KNN	82.5	85.0	81.0	81.0	92
Coarse KNN	69.5	81.5	62.0	62.0	91
Cosine KNN	85.9	86.0	85.5	85.5	94
Cubic KNN	79.1	81.0	78.0	78.0	88
Weighted KNN	82.0	85.0	80.5	80.5	93

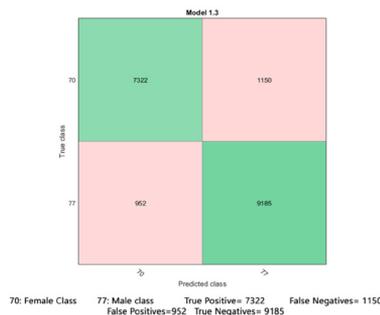


Fig. 16. Confusion Matrix on Upper body PETA dataset

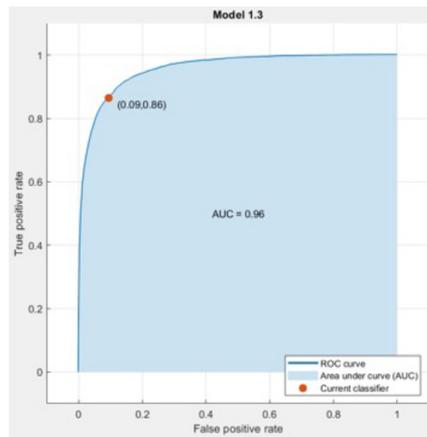


Fig. 16. ROC Curve on Upper body PETA dataset

4.1.10. Comparisons of results on Datasets

We have evaluated the algorithm on all three datasets, i.e., PETA Dataset Full body, PETA Dataset lower body, PETA Dataset upper. Best results are achieved on the upper body dataset. The comparison of the results on Cubic SVM is illustrated in Table 14.

Table 14: Result comparisons on all dataset using Cubic SVM

Dataset	Classifier	Accuracy	Precision	Sensitivity	Specificity	AUC
Full body	Cubic	84.8%	85%	84.5%	84.5%	92
Lower body	Cubic	80.0%	80%	79.0%	79.0%	92
Upper Body	Cubic	88.7%	88.5%	88.5%	88.5%	96

4.1.11. Comparisons with the related work

Result shows (Table 15) that our proposed methodology has outperformed all the existing methods such as HOG, Mini-CNN, AlexNet-CNN (Antipov, G., Berrani, S. A., et al., 2015, October), Hierarchical ELM (Zhu, W., Miao, J., et al., 2015, July), GoogleNet (Szegedy, C., Liu, W., Jia, Y., et al., 2015), ResNet50 (Chen, S., Lach, J., Lo, B., & Yang, G. Z., 2016), SSAE (Raza, M., et al., 2018), W-HOG, DFL, HDFL (Cai, L., Zhu, J., Zeng, H., et al., 2018). and DFN, DHFFN (Cai, L., Zhu, J., Zeng, H., Chen, J., & Cai, C., 2018).

Table 15: Performance comparison between the proposed method and existing methods

Author	Year	AUC	Methodology
Antipov, G., Berrani, S. A., et al. (2015, October)	2015	90	Alexnet-CNN
Antipov, G., Berrani, S. A., et al. (2015, October)	2015	86	Mini-CNN
Antipov, G., Berrani, S. A., et al. (2015, October)	2015	88	HOG
Zhu, W., Miao, J., et al. (2015, July)	2015	92	Hierarchical ELM

Szegedy, C., Liu, W., Jia, Y., et al. (2015).	2015	91	GooleNet
Chen, S., Lach, J., Lo, B., & Yang, G. Z. (2016)	2016	90	ResNet50
Raza et al. (Raza, M., et al., 2018)	2018	92	SSAE
Cai, L., Zhu, J., Zeng, H., et al. (2018)	2018	86	W-HOG
Cai, L., Zhu, J., Zeng, H., et al. (2018)	2018	93	DFL
Cai, L., Zhu, J., Zeng, H., et al. (2018)	2018	95	HDFL
Cai, L., Zhu, J., Zeng, H., Chen, J., & Cai, C. (2018)	2018	93	DFN
Cai, L., Zhu, J., Zeng, H., Chen, J., & Cai, C. (2018)	2018	95	DHFFN
Proposed	2020	96	Feature Ensembles

Results are compared based on datasets; in the past, no existing method has used the full PETA Dataset because of its diverse nature. There are so many images that pose variations in the dataset, due to which it is always been a complex task to validate and evaluate the technique in such an environment.

4.2. Discussion and analysis

We have deeply analyzed the experimental results and have also reviewed the performances of our proposed algorithm. Our results have also been compared with the existing methodologies. Our technique could be used in the future for various object activity recognition too. Generally, our introduced methodology has four steps a) feature extraction, c) feature fusion, d) classification. We have taken 5 performance measures: Accuracy, Specificity, Sensitivity, Precision, and Area under the curve. Our proposed methodology uses three feature extractors: HOG, Gabor, and Granulometry, and the best feature is fused to achieve a strong feature vector. Later we have applied classification methods such as SVM and KNN to acquire the best possible results. Results are evaluated on the PETA dataset, a combination of 10 total different pedestrian datasets and very diverse. Using the PETA dataset, we have also created two datasets on the total ratio of the human body. We have divided the whole body into 2:3 to separate the upper and lower body. We have achieved the highest classification rate with the cubic SVM classifier with an accuracy of 88.7% and 96% AUC. We have also compared the results with existing methodologies. Our results verified that the proposed technique provides good results as compared to the existing techniques.

Conclusion

In this work, we have proposed a methodology for gender pedestrian recognition, for this purpose, we have done preprocessing on the PETA dataset and have made three sets of the dataset in which we have full body, upper body, and lower body of male and female. After the preprocessing, we have conducted 5 tests by using different feature values for each feature extractor. In our method, we have used three feature extractors which are HOG, Gabor, and Granulometry. The feature vector for each feature has been carried out and is fused to acquire a strong feature vector. The fused feature is utilized for classification. To validate and evaluate our proposed technique, we used the SVM and KNN, and we can conclude that our existing

technique is significantly better compared to the existing techniques and achieve the highest accuracy of 88.7% and AUC 0.96.

Although good results are attained, there is always room for improvement. In the future, it is possible to improve the accuracy more than this by using some other feature extractor and fuse the ideal score of each feature to make a stronger feature vector. For classification, other classifiers can be used. There are also deep learning methods and techniques that can be used with various handcrafted features to achieve better results with different performance measures.

References

Agaian, S. S., Silver, B., & Panetta, K. A. (2007). Transform coefficient histogram-based image enhancement algorithms using contrast entropy. *IEEE transactions on image processing*, 16(3), 741-758.

Ali, H., Sharif, M., Yasmin, M., Rehmani, M. H., & Riaz, F. (2020). A survey of feature extraction and fusion of deep learning for detection of abnormalities in video endoscopy of gastrointestinal-tract. *Artificial Intelligence Review*, 53(4), 2635-2707.

Almudhahka, N., Nixon, M., & Hare, J. (2016, February). Human face identification via comparative soft biometrics. In *2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)* (pp. 1-6). IEEE.

Alterman, R., Zito-Wolf, R., & Carpenter, T. (1998). Pragmatic action. *Cognitive Science*, 22(1), 53-105.

Amin, J., Sharif, M., et al. (2018). Diabetic retinopathy detection and classification using hybrid feature set. *Microscopy research and technique*, 81(9), 990-996.

Amin, J., Sharif, M., et al. (2019). Brain tumor detection using statistical and machine learning method. *Computer methods and programs in biomedicine*, 177, 69-79.

Amin, J., Sharif, M., et al. (2019, April). Brain tumor classification: feature fusion. In *2019 international conference on computer and information sciences (ICCIS)* (pp. 1-6). IEEE.

Amin, J., Sharif, M., et al. (2020). An Integrated Design for Classification and Localization of Diabetic Foot Ulcer based on CNN and YOLOv2-DFU Models. *IEEE Access*.

Amin, J., Sharif, M., et al. (2020d). Convolutional neural network with batch normalization for glioma and stroke lesion detection using MRI. *Cognitive Systems Research*, 59, 304-311.

Amin, J., Sharif, M., et al. (2020e). Use of machine intelligence to conduct analysis of human brain data for detection of abnormalities in its cognitive functions. *Multimedia Tools and Applications*, 79(15), 10955-10973.

Amin, J., Sharif, M., Gul, N., Raza, M., Anjum, M. A., Nisar, M. W., & Bukhari, S. A. C. (2020). Brain tumor detection by using stacked autoencoders in deep learning. *Journal of medical systems*, 44(2), 1-12.

Amin, J., Sharif, M., Raza, M., & Yasmin, M. (2018). Detection of brain tumor based on features fusion and machine learning. *Journal of Ambient Intelligence and Humanized Computing*, 1-17.

Amin, J., Sharif, M., Raza, M., Saba, T., Sial, R., & Shad, S. A. (2020). Brain tumor detection: A long short-term memory (LSTM)-based learning model. *Neural Computing and Applications*, 32(20), 15965-15973.

Ansari, G. J., Shah, J. H., Sharif, M., & ur Rehman, S. (2020). A novel approach for

scene text extraction from synthesized hazy natural images. *Pattern Analysis and Applications*, 23(3), 1305-1322.

Antipov, G., Berrani, S. A., et al. (2015, October). Learned vs. hand-crafted features for pedestrian gender recognition. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 1263-1266).

Arshad, H., Khan, M. A., et al. (2019). Multi-level features fusion and selection for human gait recognition: an optimized framework of Bayesian model and binomial distribution. *International Journal of Machine Learning and Cybernetics*, 10(12), 3601-3618.

Ayyaz, M. N., Javed, I., & Mahmood, W. (2016). Handwritten character recognition using multiclass svm classification with hybrid feature extraction. *Pakistan Journal of Engineering and Applied Sciences*.

Azeem, A., Sharif, M., et al. (2015). Hexagonal scale invariant feature transform (H-SIFT) for facial feature extraction. *Journal of applied research and technology*, 13(3), 402-408.

Basha, A. F., et al. (2012). Face gender image classification using various wavelet transform and support vector machine with various kernels. *International Journal of Computer Science Issues (IJCSI)*, 9(6), 150.

Bekios-Calfa, J., Buenaposada, J. M., & Baumela, L. (2014). Robust gender recognition by exploiting facial attributes dependencies. *Pattern recognition letters*, 36, 228-234.

Buades, A., Coll, B., & Morel, J. M. (2005, June). A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 2, pp. 60-65). IEEE.

Buhr, J. D., Goodwin, R. M., et al. (2000). *U.S. Patent No. 6,097,470*. Washington, DC: U.S. Patent and Trademark Office.

Cai, L., Zhu, J., Zeng, H., Chen, J., & Cai, C. (2018). Deep-learned and hand-crafted features fusion network for pedestrian gender recognition. In *Proceedings of ELM-2016* (pp. 207-215). Springer, Cham.

Cai, L., Zhu, J., Zeng, H., et al. (2018). HOG-assisted deep feature learning for pedestrian gender recognition. *Journal of the Franklin Institute*, 355(4), 1991-2008.

Cao, C., Schultz, A. B., et al. (1998). Sudden turns and stops while walking: kinematic sources of age and gender differences. *Gait & Posture*, 7(1), 45-52.

Cao, L., Dikmen, M., Fu, Y., & Huang, T. S. (2008, October). Gender recognition from body. In *Proceedings of the 16th ACM international conference on Multimedia* (pp. 725-728).

Chen, S., Lach, J., Lo, B., & Yang, G. Z. (2016). Toward pervasive gait analysis with wearable sensors: A systematic review. *IEEE journal of biomedical and health informatics*, 20(6), 1521-1537.

Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE.

Derpanis, K. G. (2007). Gabor filters.

Dibra, E., Jain, H., Oztireli, C., Ziegler, R., & Gross, M. (2017). Human shape from silhouettes using generative hks descriptors and cross-modal neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4826-4836).

- Dougherty, E. (2018). *Mathematical morphology in image processing*. CRC press.
- Enzweiler, M., Eigenstetter, A., et al. (2010, June). Multi-cue pedestrian classification with partial occlusion handling. In **2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition** (pp. 990-997). IEEE.
- Esteban, A. et al. (2003). A rigorous and efficient full-wave analysis of uniform bends in rectangular waveguide under arbitrary incidence. *IEEE Transactions on microwave theory and techniques*, 51(2), 397-405.
- Fayyaz, M., Yasmin, M., et al. (2020). Person re-identification with features-based clustering and deep features. *Neural Computing and Applications*, 32(14), 10519-10540.
- Fayyaz, M., Yasmin, M., Sharif, M., & Raza, M. (2021). J-LDFR: joint low-level and deep neural network feature representations for pedestrian gender classification. *Neural Computing and Applications*, 33, 361-391.
- Ferrari, S., Piuri, V., & Scotti, F. (2008, July). Image processing for granulometry analysis via neural networks. In *2008 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications* (pp. 28-32). IEEE.
- Ge, W., Collins, R. T., & Ruback, B. (2009, December). Automatically detecting the small group structure of a crowd. In *2009 Workshop on Applications of Computer Vision (WACV)* (pp. 1-8). IEEE.
- Golomb, B. A., Lawrence, D. T., & Sejnowski, T. J. (1990, November). SEXNET: A Neural Network Identifies Sex From Human Faces. In *NIPS* (Vol. 1, p. 2).
- Gromski, P. S., Muhamadali, H., et al. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis—a marriage of convenience or a shotgun wedding. *Analytica chimica acta*, 879, 10-23.
- Guo, G., Dyer, C. R., Fu, Y., & Huang, T. S. (2009, September). Is gender recognition affected by age?. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops* (pp. 2032-2039). IEEE.
- Hadid et al. (2008). An experimental comparison of gender classification methods. *pattern recognition letters*, 29(10), 1544-1556.
- Hadid, A., & Pietikäinen, M. (2009). Combining appearance and motion for face and gender recognition from videos. *Pattern Recognition*, 42(11), 2818-2827.
- Hamamoto, Y., Uchimura, S., et al. (1998). A Gabor filter-based method for recognizing handwritten numerals. *Pattern recognition*, 31(4), 395-400.
- Herweg, J. A., Kerekes, J. P., Ientilucci, E. J., & Eismann, M. T. (2011, June). Spectral variations in HSI signatures of thin fabrics for detecting and tracking of pedestrians. In *Active and Passive Signatures II* (Vol. 8040, p. 80400G). International Society for Optics and Photonics.
- Hussain, S. J., Chen, Z., et al. (2016, January). Color Based Pre-rank Categorization for Person Re-identification. In *2016 International Conference on Intelligent Control and Computer Application (ICCA 2016)*. Atlantis Press.
- Irum, I., Shahid, M. A., Sharif, M., & Raza, M. (2015). A Review of Image Denoising Methods. *Journal of Engineering Science & Technology Review*, 8(5).
- Jahromi, M. N., Bonderup, M. B., et al. (2018, March). Automatic access control based on face and hand biometrics in a non-cooperative context. In *2018 IEEE Winter Applications of Computer Vision Workshops (WACVW)* (pp. 28-36). IEEE.
- Keerthi, S. S., Shevade, S. K., et al. (2001). Improvements to Platt's SMO algorithm for

SVM classifier design. *Neural computation*, 13(3), 637-649.

Khan, M. A., Akram, T., Sharif, M., Javed, K., Raza, M., & Saba, T. (2020). An automated system for cucumber leaf diseased spot detection and classification using improved saliency method and deep features selection. *Multimedia Tools and Applications*, 1-30.

Khan, M. A., et al. (2019). An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection. *Neural Computing and Applications*, 1-20.

Khan, M. A., Kadry, S., et al. (2021). Prediction of COVID-19-pneumonia based on selected deep features and one class kernel extreme learning machine. *Computers & Electrical Engineering*, 90, 106960.

Khan, M. A., Sharif, M. I., Raza, M., Anjum, A., Saba, T., & Shad, S. A. (2019). Skin lesion segmentation and classification: A unified framework of deep neural network features fusion and selection. *Expert Systems*, e12497.

Khan, M. A., Sharif, M., et al. (2020). Hand-crafted and deep convolutional neural network features fusion and selection strategy: an application to intelligent human action recognition. *Applied Soft Computing*, 87, 105986.

Khan, M. A., Zhang, Y. D., Sharif, M., & Akram, T. (2021). Pixels to classes: intelligent learning framework for multiclass skin lesion localization and classification. *Computers & Electrical Engineering*, 90, 106956.

Li, B., Lian, X. C., & Lu, B. L. (2012). Gender classification by combining clothing, hair and facial component classifiers. *Neurocomputing*, 76(1), 18-27.

Li, D., Zhang, Z., Chen, X., et al. (2016). A richly annotated dataset for pedestrian attribute recognition. *arXiv preprint arXiv:1603.07054*.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.

Liu, C., & Wechsler, H. (2002). Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image processing*, 11(4), 467-476.

M. Simonelli and A. Quaglio (2015). Surveillance camera, ed: Google Patents.

MageshKumar, C., Thiyagarajan, R., et al. (2011, March). Gabor features and LDA based face recognition with ANN classifier. In *2011 International Conference on Emerging Trends in Electrical and Computer Technology* (pp. 831-836). IEEE.

Masood, S., Sharif, M., Raza, M., Yasmin, M., Iqbal, M., & Younus Javed, M. (2015). Glaucoma disease: A survey. *Current Medical Imaging*, 11(4), 272-283.

Mathanker, S. K., Weckler, P. R., et al. (2011). AdaBoost classifiers for pecan defect classification. *Computers and electronics in agriculture*, 77(1), 60-68.

Meyer, D., & Wien, F. T. (2015). Support vector machines. *The Interface to libsvm in package e1071*, 28.

Munir, A., Hussain, A., et al. (2018). Illumination invariant facial expression recognition using selected merged binary patterns for real world images. *Optik*, 158, 1016-1025.

Naz, J., Sharif, M., Yasmin, M., Raza, M., & Khan, M. A. (2020). Detection and Classification of Gastrointestinal Diseases using Machine Learning. *Current Medical Imaging*.

Ng, C. B., Tay, Y. H., & Goi, B. M. (2012). Vision-based human gender recognition: A survey. *arXiv preprint arXiv:1204.1611*.

Papanikolopoulos, N. P., Krantz, D. G., et al. (2003). *U.S. Patent No. 6,548,982*. Wash-

ington, DC: U.S. Patent and Trademark Office.

Persoon, E., & Fu, K. S. (1977). Shape discrimination using Fourier descriptors. *IEEE Transactions on systems, man, and cybernetics*, 7(3), 170-179.

Piccardi, M. (2004, October). Background subtraction techniques: a review. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)* (Vol. 4, pp. 3099-3104). IEEE.

Prodanov, D., Heeroma, J., & Marani, E. (2006). Automatic morphometry of synaptic boutons of cultured cells using granulometric analysis of digital images. *Journal of neuroscience methods*, 151(2), 168-177.

Qureshi, I., Sharif, M., et al. (2016). Computer aided systems for diabetic retinopathy detection using digital fundus images: A survey. *Current Medical Imaging*, 12(4), 234-241.

R. L. Tucker (2000), Sediment filtering system, ed: Google Patents.

Rashid, M., Khan, M. A., et al. (2019). Object detection and classification: a joint selection and fusion strategy of deep convolutional neural network and SIFT point features. *Multimedia Tools and Applications*, 78(12), 15751-15777.

Rauf, H. T., Saleem, B. A., et al. (2019). A citrus fruits and leaves dataset for detection and classification of citrus diseases through machine learning. *Data in brief*, 26, 104340.

Raza, M., Chen, Z., et al. (2018). Framework for estimating distance and dimension attributes of pedestrians in real-time environments using monocular camera. *Neurocomputing*, 275, 533-545.

Raza, M., et al. (2018). Appearance based pedestrians' gender recognition by employing stacked auto encoders in deep learning. *Future Generation Computer Systems*, 88, 28-39.

Raza, M., Sharif, M., et al. (2018). Appearance based pedestrians' gender recognition by employing stacked auto encoders in deep learning. *Future Generation Computer Systems*, 88, 28-39.

Raza, M., Zonghai, C., et al. (2017). Part-wise pedestrian gender recognition via deep convolutional neural networks.

Raza, M., Zonghai, C., et al. (2017, August). Pedestrian classification by using stacked sparse autoencoders. In *2017 2nd International Conference on Advanced Robotics and Mechatronics (ICARM)* (pp. 37-42). IEEE.

Rehman, S. U., Chen, Z., et al. (2018). Person re-identification post-rank optimization via hypergraph-based learning. *Neurocomputing*, 287, 143-153.

Rehman, S., Chen, Z., et al. (2016, July). Multi-feature fusion based re-ranking for person re-identification. In *2016 International Conference on Audio, Language and Image Processing (ICALIP)* (pp. 213-216). IEEE.

Ross, L., & Russ, J. C. (2011). The image processing handbook. *Microscopy and Microanalysis*, 17(5), 843.

Russell, S., & Norvig, P. (2002). Artificial intelligence: a modern approach.

Saba, T., Bokhari, S. T. F., et al. (2018). Fundus image classification methods for the detection of glaucoma: A review. *Microscopy research and technique*, 81(10), 1105-1121.

Saba, T., et al. (2019). Lung nodule detection based on ensemble of hand crafted and deep features. *Journal of medical systems*, 43(12), 1-12.

Saba, T., Rehman, A., Jamail, N. S. M., et al. (2021). Categorizing the Students' Activities for Automated Exam Proctoring Using Proposed Deep L2-GraftNet CNN Network and ASO Based Feature Selection Approach. *IEEE Access*, 9, 47639-47656.

Schumann, A., & Stiefelhagen, R. (2017). Person re-identification by deep learning attribute-complementary information. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 20-28).

Seo, C., Lee, K. Y., & Lee, J. (2001). GMM based on local PCA for speaker identification. *Electronics Letters*, 37(24), 1486-1488.

Shah, G. A., Khan, A., et al. (2015). A Review on Image Contrast Enhancement Techniques using Histogram Equalization. *Science International*, 27(2).

Shan, C. (2012). Learning local binary patterns for gender classification on real-world face images. *Pattern recognition letters*, 33(4), 431-437.

Sharif, M. I., Khan, M. A., Alhussein, M., Aurangzeb, K., & Raza, M. (2021). A decision support system for multimodal brain tumor classification using deep learning. *Complex & Intelligent Systems*, 1-14.

Sharif, M., Amin, J., et al. (2020a). Recognition of Different Types of Leukocytes Using YOLOv2 and Optimized Bag-of-Features. *IEEE Access*, 8, 167448-167459.

Sharif, M., Amin, J., et al. (2020b). Brain tumor detection based on extreme learning. *Neural Computing and Applications*, 1-13.

Sharif, M., Amin, J., et al. (2020c). An integrated design of particle swarm optimization (PSO) with fusion of features for detection of brain tumor. *Pattern Recognition Letters*, 129, 150-157.

Sharif, M., Irum, I., Yasmin, M., & Raza, M. (2017). Salt & Pepper Noise Removal from Digital Color Images Based on Mathematical Morphology and Fuzzy Decision. *Nepal Journal of Science and Technology*, 18(1), 1-7.

Sharif, M., Khan, S., et al. (2019, April). Improved Video Stabilization using SIFT-Log Polar Technique for Unmanned Aerial Vehicles. In *2019 International Conference on Computer and Information Sciences (ICIS)* (pp. 1-7). IEEE.

Sharif, M., Raza, M., et al. (2019). An overview of biometrics methods. *Handbook of Multimedia Information Security: Techniques and Applications*, 15-35.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8), 888-905.

Struijk, L., Bavinck, J. N. B., et al. (2003). Presence of human papillomavirus DNA in plucked eyebrow hairs is associated with a history of cutaneous squamous cell carcinoma. *Journal of Investigative Dermatology*, 121(6), 1531-1535.

Szegedy, C., Liu, W., Jia, Y., et al. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

Ueki, K., Komatsu, H., Imaizumi, S., et al. (2004, August). A method of gender classification by integrating facial, hairstyle, and clothing images. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.* (Vol. 4, pp. 446-449). IEEE.

Vincent, L. (1994). Fast grayscale granulometry algorithms. In *Mathematical morphology and its applications to image processing* (pp. 265-272). Springer, Dordrecht.

Wen, X., Shao, L., et al. (2015). A rapid learning algorithm for vehicle classification. *Information sciences*, 295, 395-406.

Williams, L. M., Mathersul, D., et al. (2009). Explicit identification and implicit recognition of facial emotions: I. Age effects in males and females across 10 decades. *Journal of Clinical and Experimental Neuropsychology*, 31(3), 257-277.

Won, A. S., Yu, L., et al. (2012). Tracking gesture to detect gender. In *Proc. of the International Society for Presence Research Annual Conference* (pp. 24-26).

Xia, B., et al. (2013, September). Enhancing gender classification by combining 3D and 2D face modalities. In *21st European Signal Processing Conference (EUSIPCO 2013)* (pp. 1-5). IEEE.

Xu, J., Liu, J., Yin, J., & Sun, C. (2016). A multi-label feature extraction algorithm via maximizing feature variance and feature-label dependence simultaneously. *Knowledge-Based Systems*, 98, 172-184.

Yang, J., Liu, L., Jiang, T., & Fan, Y. (2003). A modified Gabor filter design method for fingerprint image enhancement. *Pattern Recognition Letters*, 24(12), 1805-1817.

Yu, Z., Chen, H., Liu, J., et al. (2015). Hybrid k-nearest neighbor classifier. *IEEE transactions on cybernetics*, 46(6), 1263-1275.

Yuan et al. (2016). Human gender classification: a review. *International Journal of Biometrics*, 8(3-4), 275-300.

Zhang, Z. (2012). Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2), 4-10.

Zhu, W., Miao, J., et al. (2015, July). Hierarchical extreme learning machine for unsupervised representation learning. In *2015 international joint conference on neural networks (ijcnn)* (pp. 1-8). IEEE.

Ziems, M., Breitkopf, U., Heipke, C., & Rottensteiner, F. (2012, August). Multiple-model based verification of road data. In *XXII ISPRS Congress* (Vol. 25).

Submitted: 02.07.2020

Accepted: 21.05.2021