Krizhevsky et al. [8] presented the AlexNet in the ImageNet LSVRC-2010 contest. In this contest, this architecture gave quite successful result to classify over 1 million images into 1000 classes. AlexNet comprises five convolution layer, three pooling layers and three fully connected layer respectively. In this architecture, the final convolutional feature map is formed as a vector to become an input to the fully connected layers. The image descriptor generated by AlexNet is represented by the output of this layer.

The face recognition application designed here with deep learning methods requires a pre-processing stage which includes modifications on images. This stage is an important because it can overcome the variation effects in the images such as illumination, occlusion, the background of the image. The AlexNet was trained on 227x227 pixel RGB images. For this purpose, all images among the Faces95 and Faces 96 datasets was resized to required size. Also grey colored images in the AT&T dataset were converted to RGB format [9].

Next step is face detection stage and all faces was detected using Viola&Jones algorithm [10]. After the detection phase images was selected randomly either for train or for test. For this selection process, %80-%20 ratio for train and test was used.

After the face detection stage, feature extraction was achieved using CNN which based on AlexNet. The mentioned network architecture consists of 8 main layers, 5 of which are convolution and 3 of which are fully connected layers. As mentioned above, 227x227x3 pixels colored image is used as an input for the first convolutional layer. This layer consists of 96 filters with the size of 11x11x3, 3x3 stride size and zero padding. Here, 11s and 3 represent the weight and height value and depth of the filter, respectively. Stride is the number of the scanning pixel size of the filter. Padding ensures the input and output volume is the same. Two of them are the hyperparameters of the convolutional layer. In the following layer the depth value must be the same as filter size of the former layer. Result of the first convolutional layer, the size of 55x55x96 output is obtained. After non-linear function and the size of the 3x3 max pooling layer with 2x2 stride and 0 padding, the size of the 27x27x96 feature map is obtained and sends as an input for the second convolutional layer. This layer consists of 256 filters with the size of 5x5x48 with 1x1 stride and padding. This layer is followed with 3x3 max pooling layer and the size of 13x13x 256 feature map is obtained and sent third convolutional layer. This layer consists of 384 filters with the size of 3x3x256. After that 13x3x384 feature map is obtained and is sent as an input for the fourth convolutional layer. This layer consists of 384 filter with the size of 3x3x192 with 1x1 stride and 1 padding. After non-linear function, feature map is sent to the fifth convolutional layer with the 13x13x384 output size. The fifth and last convolutional layer consists of 256 filter with the size of 3x3x192 with1x1 stride and 1 padding. Non-linear function and 3x3 max pooling layer are followed fifth convolutional layer with 1x1 stride and 1 padding. After pooling layer, the size of 6x6x256 output is obtained. This output is sent to two fully connected layer is an input. After first and second fully connected layer, 0.5 rate of dropout is applied to feature map. This feature map is sent to SVM classifier for recognition.

The extracted features were used to train the SVM [11] classifier which can be seen in fig. 1. Most basic image features like edges or corners were extracted in the very first layer of this network. Deeper layers of the network generate high-level features using previous relatively low-level features. The generated high-level features are used for classification.

As mentioned before, AlexNet has been trained to classify images of some objects to the 1000 classes which cannot be used to recognize the faces. For this reason, the classification layer of this architecture was replaced with mutli-class SVM classifier. The generated outputs from last fully connected layer were employed as input to train the SVM classifier. In proposed method, one vs all approach SVM is used. In this approach, a number of SVMs are trained for a number of each individuals and each SVM separates a single class from all remaining classes [12]. Trained SVM is used to classify the test images and a result was generated using accuracy of classifier.

In this study, we designed and tested an CNN based face recognition model. The presented network is based on pre-trained AlexNet architecture. The classification layer of this model was replaced with a multi-class SVM classifier to recognize faces among the datasets namely AT&T, Faces95, and Faces96.

III. EXPERIMENTAL RESULTS AND COMPARISON

In this study, three well known and comprehensive datasets namely “AT&T” [13], “Faces95” [14] and “Faces96” [15] were used to test the system. AT&T dataset consists of 400 images of 40 different individuals. Each of image size is 92x112 pixels. Faces95 dataset consists of 1440 images of 72 different individual. Each of image size is 180x200 pixels. Faces96 dataset consists of 3040 images of 152 different individual. Each of image size is 192x192 pixels.

Fig. 2. Example images from Faces95 dataset [13]

Fig. 3.Example images from Faces95 dataset [14]

Fig. 4. Example images from Faces96 dataset [15]