Name: PRATHMESH R. KUMBHARE

CS6001 Assignment 6: Classification

REPORT

A. OUTLINE FOR PROGRAM ALGORITHM:

The algorithm for implementing the program is given below in steps:

1. Matlab script files 'MODEL_MAIN.m' are created for each model separately. The program then executes the corresponding model. Predicted values of the class for each model are plotted to provide some visualization aid.

2. Factors which can be varied to get different results for a model are:
   - Variance of Prior ('Var_Prior') for all Bayesian approach models.
   - Initialization of 'phi' or 'psi'.
   - Value of 'LAMBDA' in kernel and relevance vector models.
   - Degrees of freedom ('nu') and 'threshold' in the relevance vector model.

3. The program firstly creates a matrix 'X' for training images by looping through each file in the sub directory of training images. Inside the loop, the image data is converted into column vector and a matrix consisting column vectors of all images is saved. Also, another column matrix 'w' is created which consists of class of images. (zero for bg & one for face) Similarly the matrices 'X_test' and 'w_test' are created for the testing images. For fast and accurate results, all images are scaled down by a factor of 4 before reshaping into columns.

4. Next, the variable 'initial_phi' or 'initial_psi' (according to the model type) is initialized. For the analysis in this report, value of this variable is considered as zero. But, this can be varied from '0 to 0.09' to get slightly different results.

5. Another variable available is the variance of prior 'Var_Prior'. For selection of this parameter, firstly variance of training image matrix is computed ('sig'). Then, the value is selected so that the ratio 'Var_Prior/sig' is equal to 10. This can be changed to get different results. Similarly, the values of 'nu', 'LAMBDA' & 'threshold' are taken according to the appropriate model.

6. The function 'fit_model.m' (like 'fit_logr.m' or 'fit_klogr.m') is called by passing the required parameters. This function file in return, calls another function file 'fit_model_cost.m'. (like 'fit_logr_cost.m' or 'fit_klogr_cost.m') The cost function file finds the value of 'phi' or 'psi'

when local minimum for the cost function is achieved and then from this value the fit model file returns prediction values of class. In the relevance vector model, ‘relevant_points’ is returned along with ‘predictions’.

7. For inference algorithm, the program firstly decides whether the predicted value is face or background. (For values above 0.5, it is classified as face. Otherwise it is classified as background) This is saved in variable ‘y_cap’. Then, the program calculates following four values as results:
   - ‘Miss_Detection’ (Average difference between actual & predicted face class values)
   - ‘Miss_Detection_Num’ (Wrongly classified face image % of total face images)
   - ‘False_Alarm’ (Average difference between actual & predicted bg class values)
   - ‘False_Alarm_Num’ (Wrongly classified bg image % of total bg images)

8. Three visualization plots are created to make further analysis on the models and compare them with each other.

B. RESULTS:

This is the table for all the models implemented for **initial_phi / initial_psi = 0**:

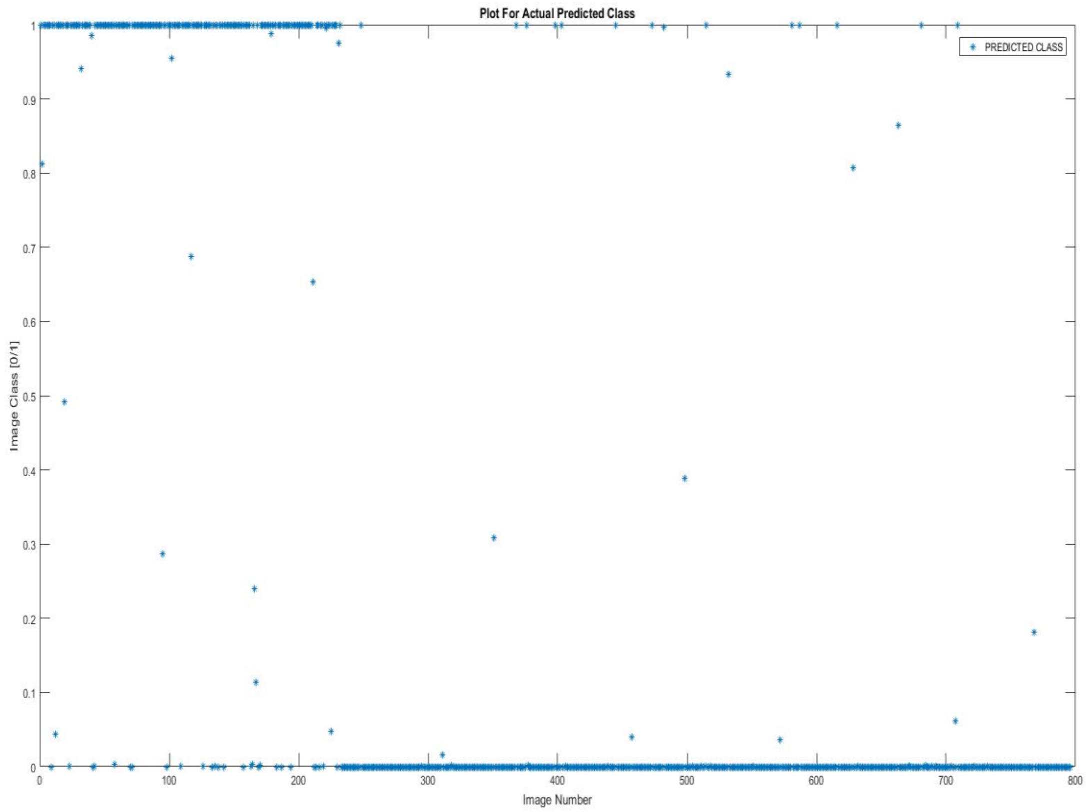
No:	MODEL	σ^2_P / σ^2_X	Run Time (sec)	MD	MD_Num (%)	FA	FA_Num (%)
1	Logistic Reg	-	20.888	0.1413	14.22	0.0313	3.01
2	Bay Logistic Reg	10	18.595	0.4985	12.93	0.4952	2.83
3	Dual Logistic Reg	-	23.431	0.1301	9.913	0.0670	3.19
4	Dual Bay Logistic Reg	10	24.211	0.4147	9.91	0.3082	3.36
5	Kernel Logistic Reg (λ=0.7)	10	22.635	0.4987	9.05	0.4966	10.10
6	Relevance Vector Logistic Reg (λ=0.7, nu=3, ν=2.62)	-	98.717	0.5000	8.62	0.4987	3.37
7	Relevance Vector Logistic Reg (Grad Img)	-	120.514				

Visualization and some information for each model is given below separately.

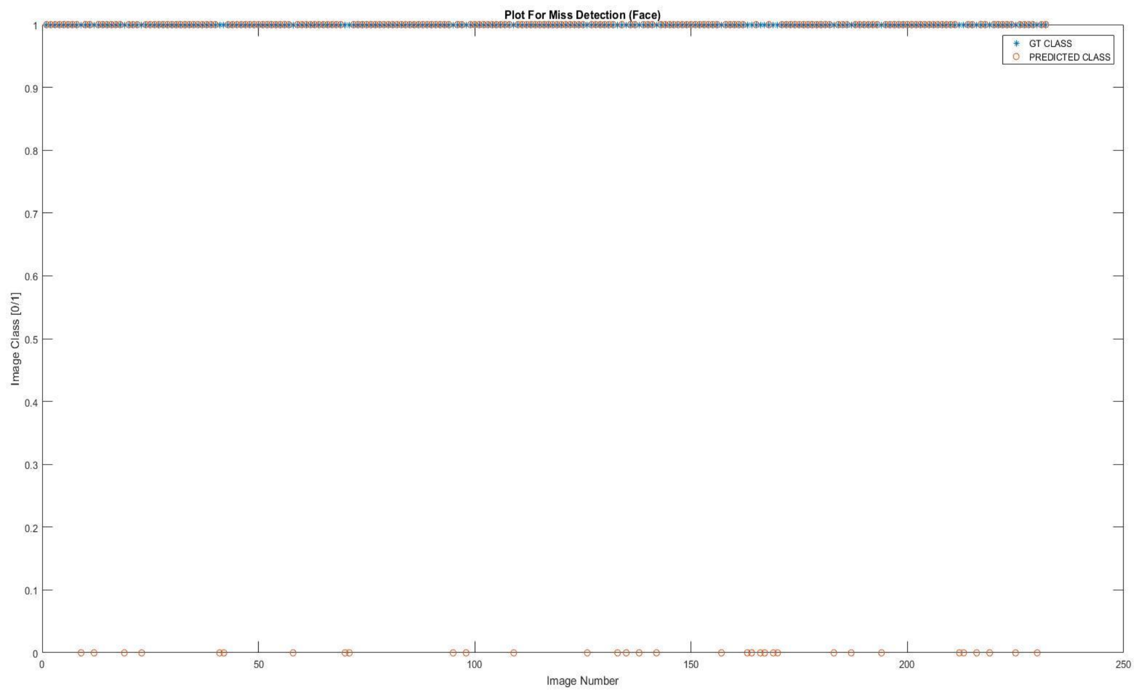
1. **Logistic Regression Model:**

- In this model, prior knowledge is not considered.
- The cost function is initialized at zero. (L=0, g=0 & H=0)
- We can see from images given below that this model is overconfident in predictions.

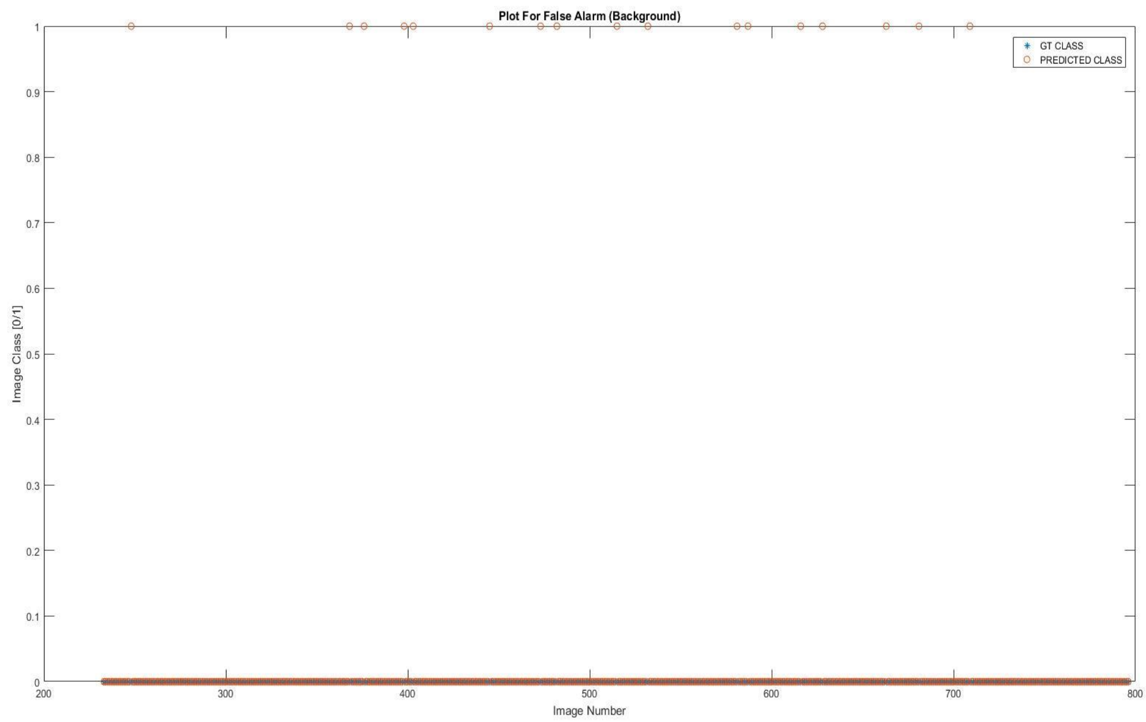
(Most of the predictions are either one or zero)



(All the circles on the x-axis are the wrongly classified face images)



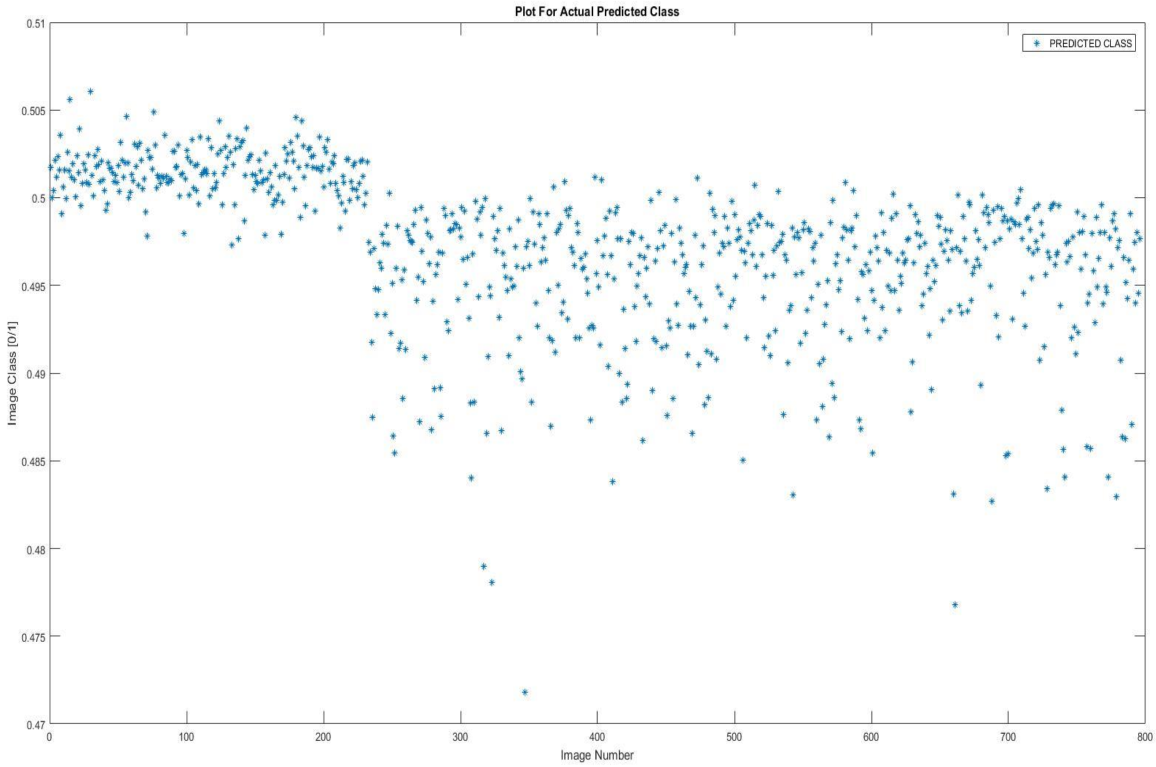
(All the circles on the upper line of the image are the wrongly classified bg images)



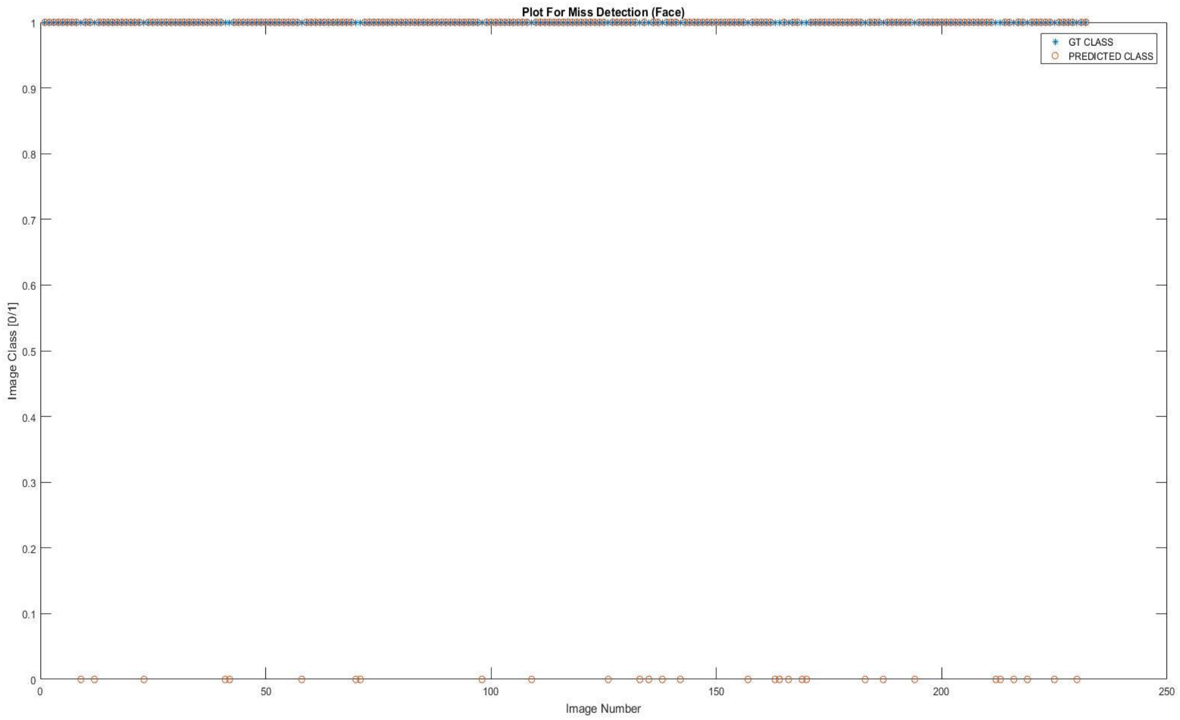
2. Bayesian Regression Model:

- In this model, prior knowledge is considered.
- Value of prior selected for this result in 10 times the variance of training image matrix.
- The cost function is initialized at some non-zero value. (Depending on prior value)
- We can see from images given below that this model is moderately confident in predictions.

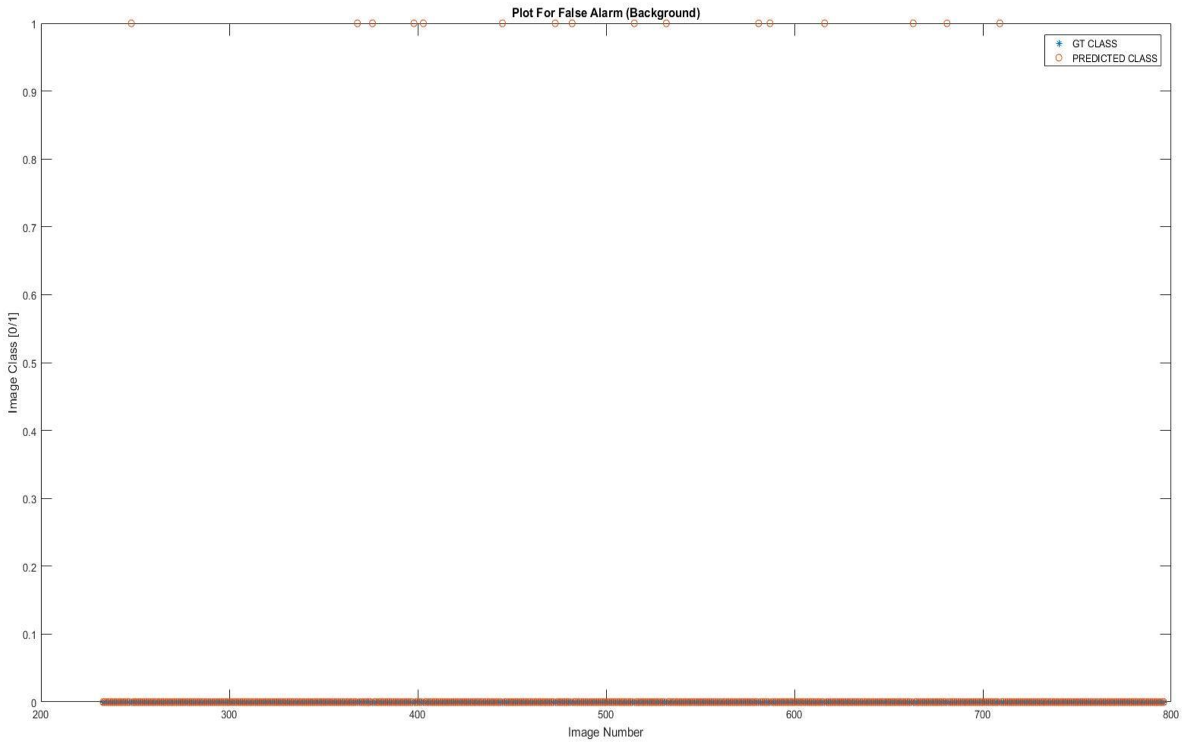
(All of the predictions are between one and zero)



(All the circles on the x-axis are the wrongly classified face images)



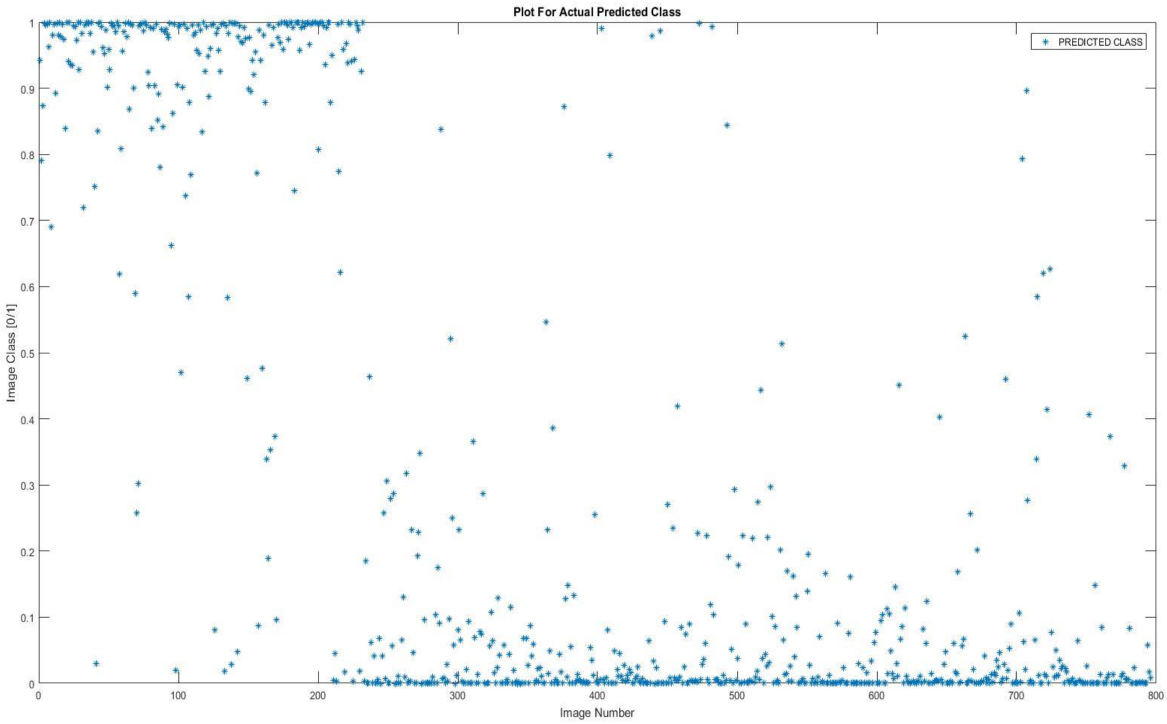
(All the circles on the upper line of the image are the wrongly classified bg images)



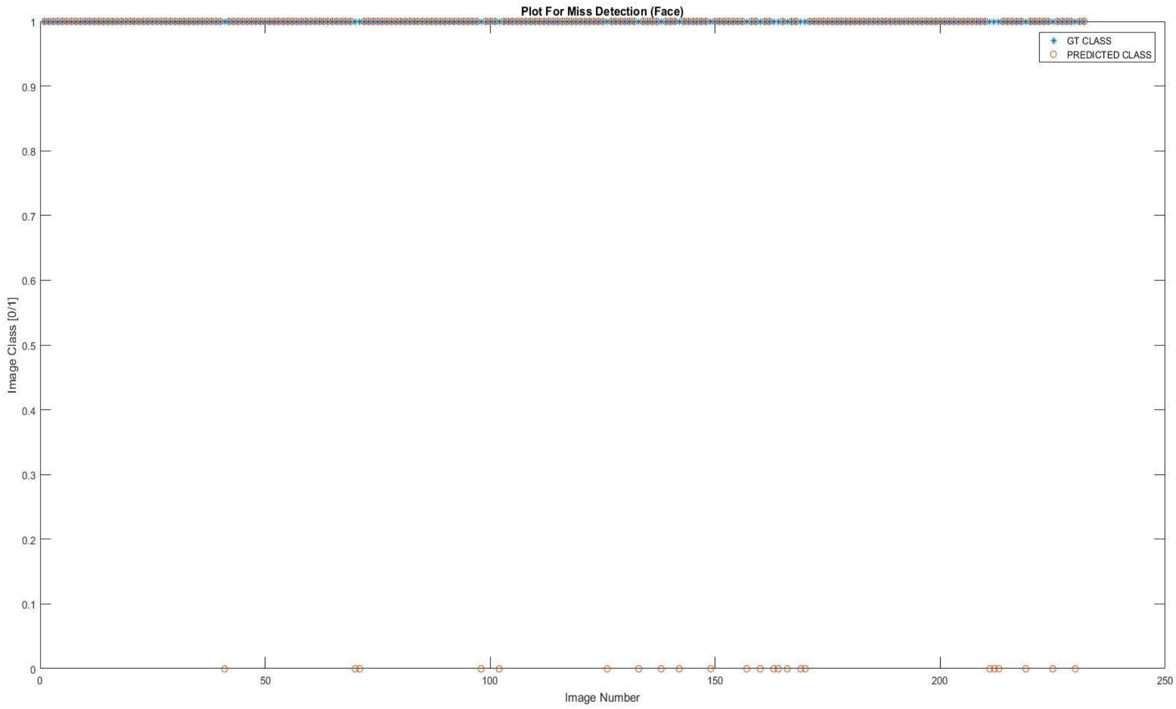
3. Dual Logistic Regression Model:

- In this model, prior knowledge is not considered.
- The cost function is initialized at zero. (L=0, g=0 & H=0)
- We can see from images given below that this model is overconfident in predictions.
- Advantage of this model is that this model computes almost similar results to logistic regression, but with **considerably less time.** From table with we see that time taken by this model is **more** than logistic model only because the image is scaled down 4 times. This effect can be seen when we are operating with images at **full scale**.

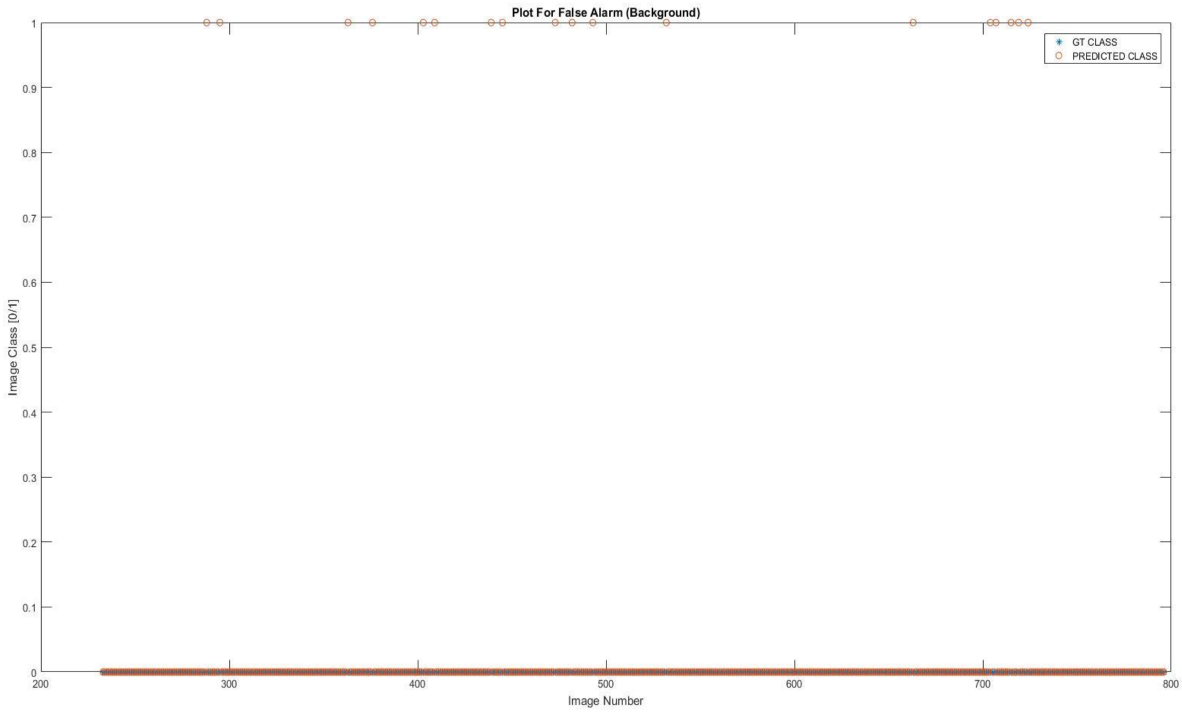
(Most of the predictions are either one or zero, but not as extreme as logistic regression model)



(All the circles on the x-axis are the wrongly classified face images)



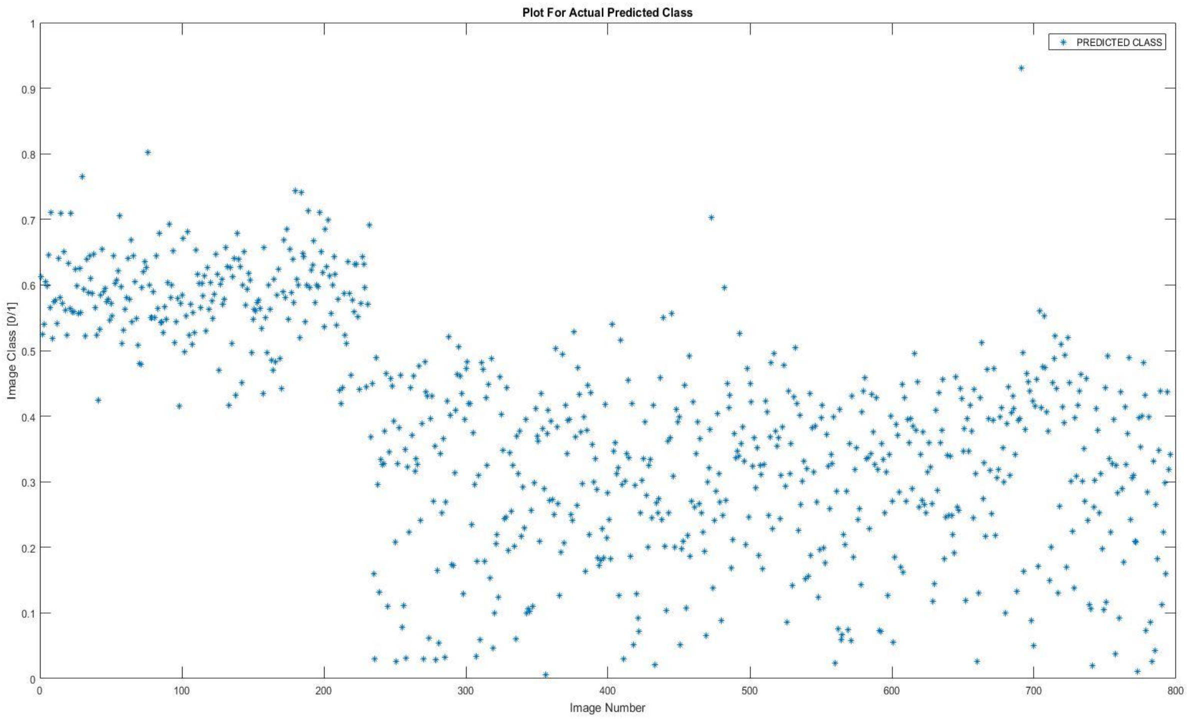
(All the circles on the upper line of the image are the wrongly classified bg images)



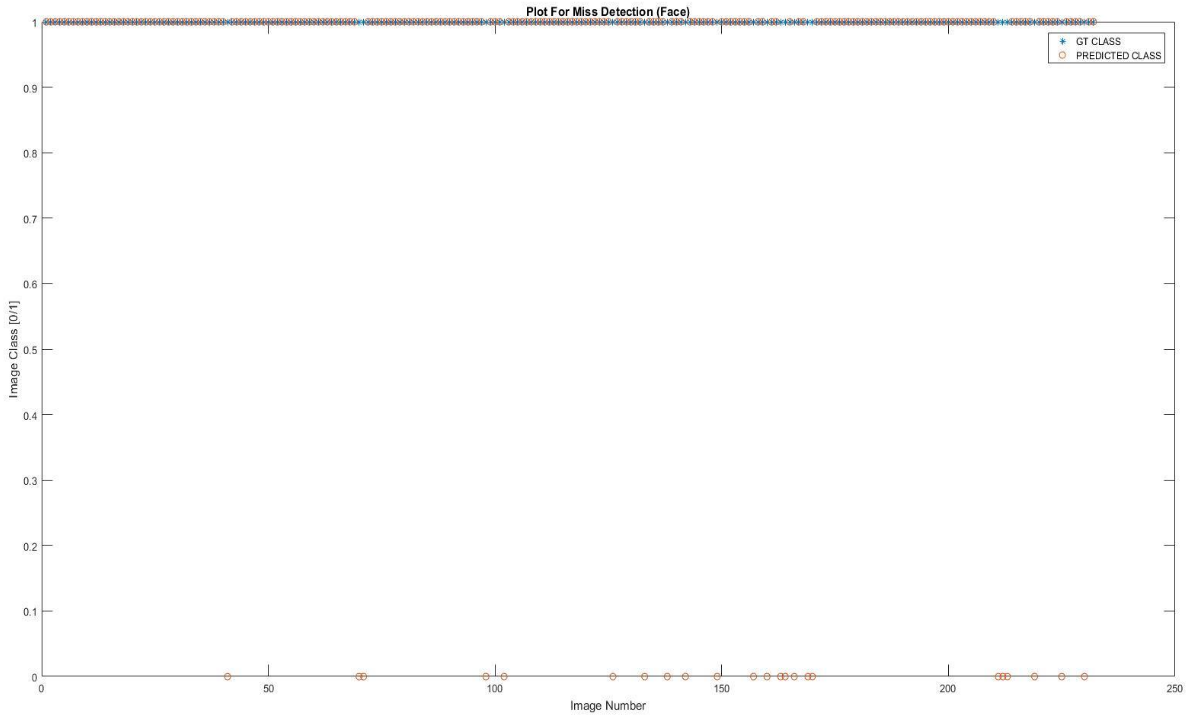
4. Dual Bayesian Regression Model:

- In this model, prior knowledge is considered.
- Value of prior selected for this result in 10 times the variance of training image matrix.
- The cost function is initialized at some non-zero value. (Depending on prior value)
- We can see from images given below that this model is moderately confident in predictions.
- Advantage of this model is that this model computes almost similar results to bayesian logistic regression, but with **considerably less time.** From table with we see that time taken by this model is **more** than bayesian logistic model only because the image is scaled down 4 times. This effect can be seen when we are operating with images at **full scale**.

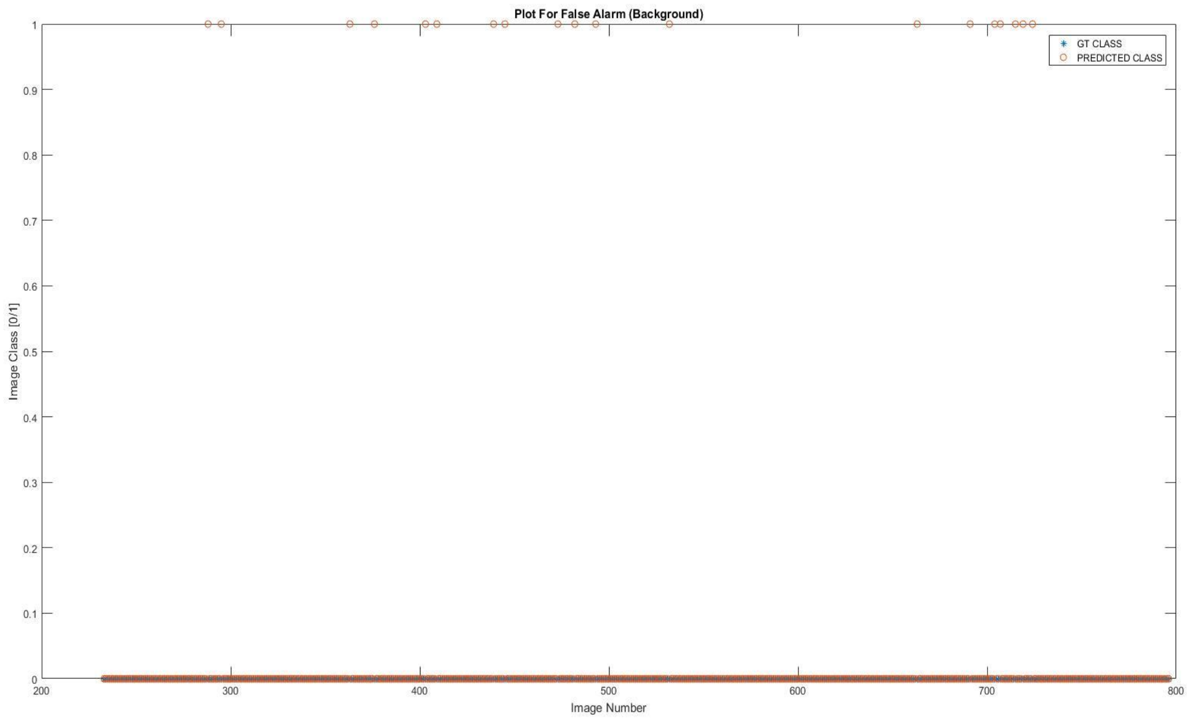
(Most the predictions are between one and zero)



(All the circles on the x-axis are the wrongly classified face images)



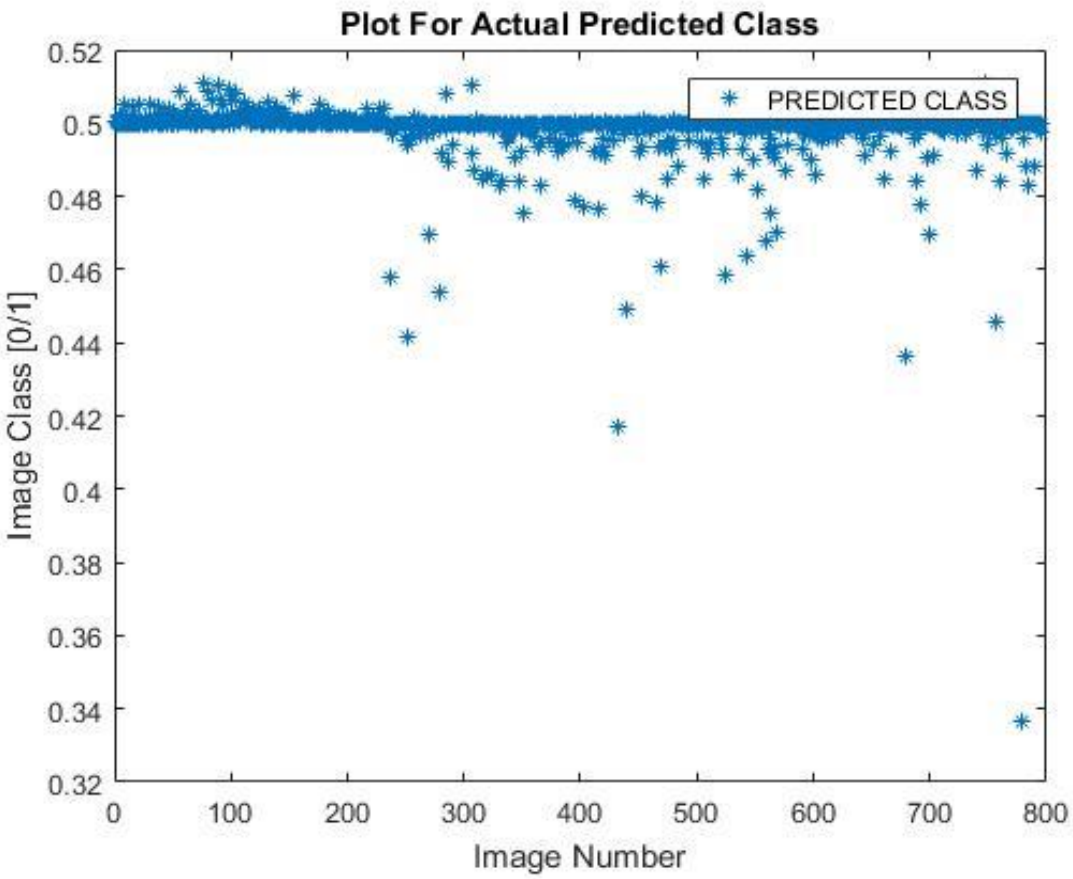
(All the circles on the upper line of the image are the wrongly classified bg images)



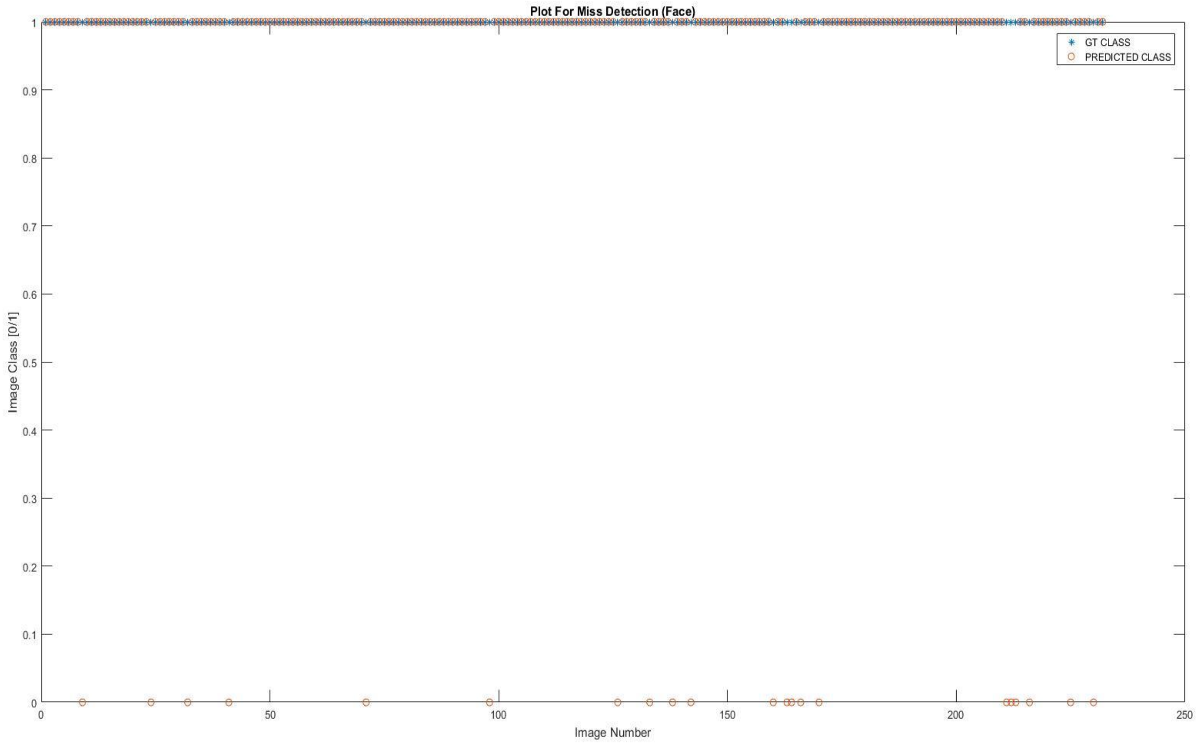
5. Kernel Logistic Regression Model:

- In this model, prior knowledge is considered.
- Value of prior selected for this result in 10 times the variance of training image matrix.
- The cost function is initialized at some non-zero value.
- Gaussian kernel function is used for computation. The λ parameter defines the **effective area** in which difference between two images is considered significant. (Like variance between two images) Generally, value in the interval [0,1] is considered. For this result value of λ=0.7 is taken.
- We can see from images given below that this model is less confident in predictions.

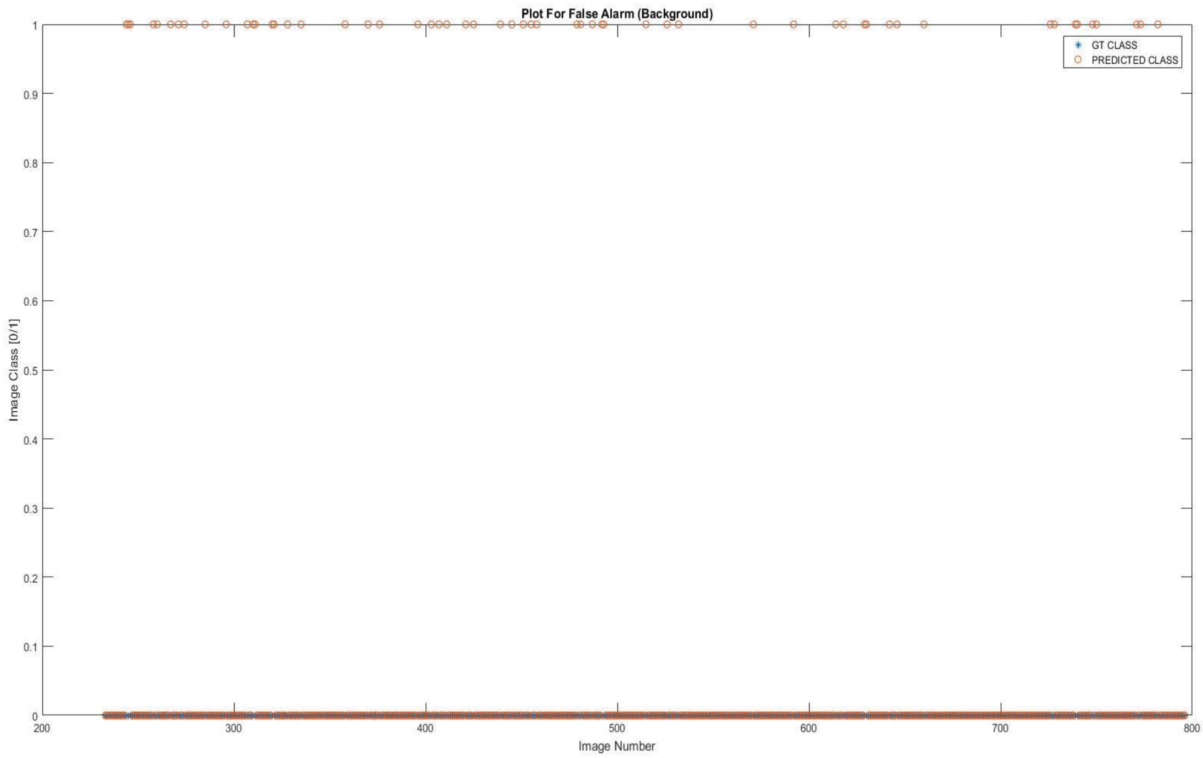
(Most the predictions are around boundary decision)



(All the circles on the x-axis are the wrongly classified face images)



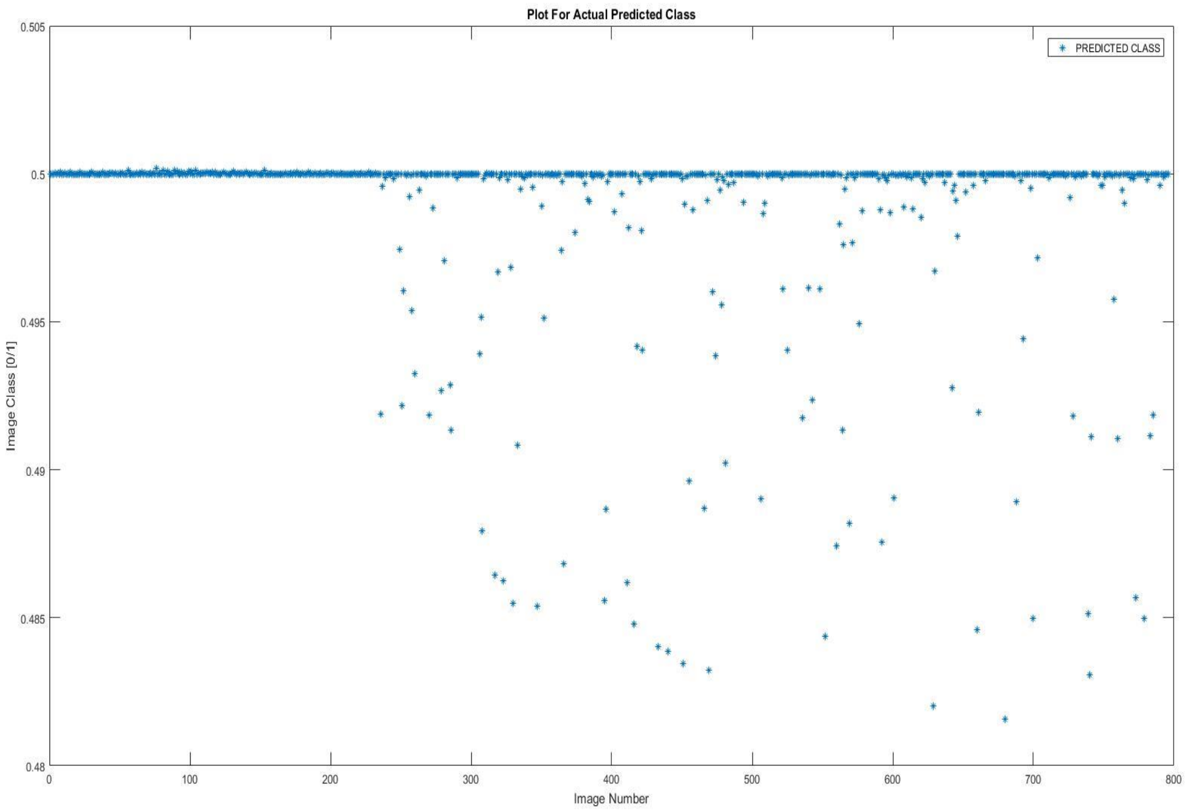
(All the circles on the upper line of the image are the wrongly classified bg images)



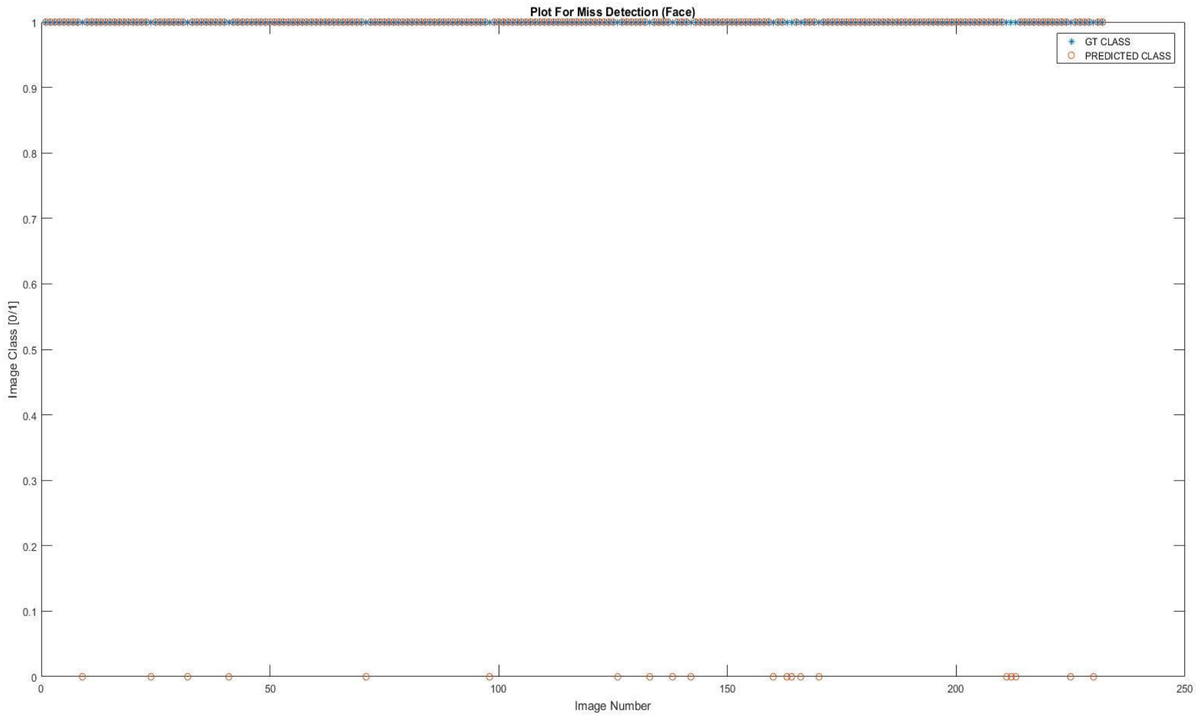
6. **Relevance Vector Logistic Regression Model:**

- In this model, prior knowledge is not considered.
- The cost function is initialized at some non-zero value.
- Gaussian kernel function is used for computation. The λ parameter defines the **effective area** in which difference between two images is considered significant. (Like variance between two images) Generally, value in the interval [0,1] is considered. For this result value of λ=0.7 is taken.
- The degrees of freedom 'nu' is the parameter used to define **amount of weight in the tails** of the t-distribution curve. Value between interval [0,2] must be avoided. As the value gets bigger, the curve approaches normal distribution. For this result value of nu=3 is taken.
- The threshold parameter ν determines **number of relevant points considered for computation of prediction.** For this result, maximum relevant points are considered. (433 in this case). But with less number of relevant points, acceptable results can be achieved.
- We can see from images given below that this model is less confident in predictions.

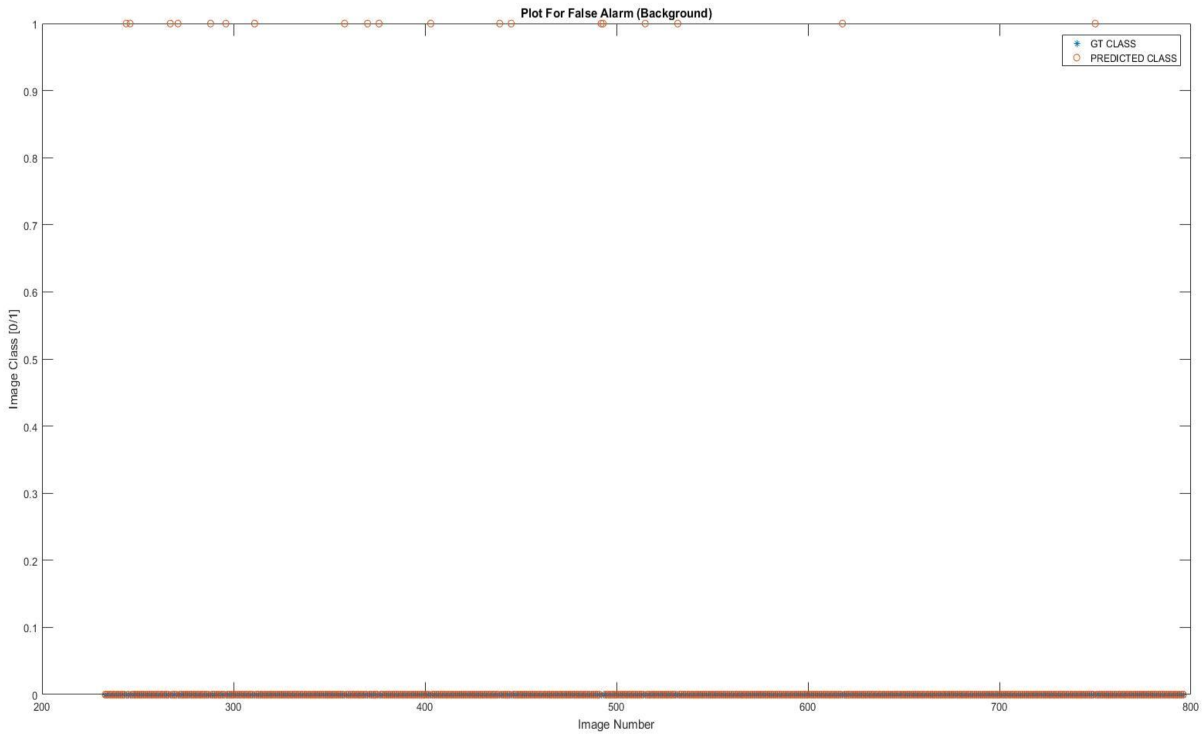
(Most the predictions are around boundary decision)



(All the circles on the x-axis are the wrongly classified face images)



(All the circles on the upper line of the image are the wrongly classified bg images)



C. General Observation & Comments:

1. The two models: logistic regression and dual logistic regression are overconfident in their predictions. Small values of **'MD'** and **'FA'** shows that these models are very certain in their predictions.
2. Similarly, the models: bayesian logistic regression and dual bayesian logistic regression are moderately confident in their predictions. Relatively large values of **'MD'** and **'FA'** shows that these models are not certain in their predictions.
3. By same pattern, the models: kernel and relevance vector logistic regression are also not certain in their predictions.
4. From the values **'MD_Num'** and **'FA_Num'** of table 1 we can see that dual models give great results considering computation time. The best results in terms of accuracy is given by the relevance vector logistic regression model.

D. Summary and Concluding Comments:

1. The aim of this assignment was to classify whether given image was face or background. From the results, it can be concluded that the algorithms classify the images with acceptable level of accuracy.
2. It could be possible to increase the accuracy of the algorithm by trying to compute in different color spaces or in grey scale to reduce computation time further.
3. One parameter which can be varied to get significantly different results is the initialization of 'initial_phi' or 'initial_psi'. The algorithm finds local minimum of the cost function, starting from zero. It may be possible that for different initialization, even better local minimum can be achieved.