CERTIFICATE

It is certified that the work contained in the thesis titled ***"Robust SVM Variants and Their Applications"*** by ***Manisha Singla*** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all requirements of Comprehensive Examination, Candidacy, and SOTA for the award of Ph.D. Degree.

Supervisor

Prof. K K Shukla

Professor,

Department of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi,
Uttar Pradesh, INDIA 221005.

Co-Supervisor

Dr. Debdas Ghosh

Assistant Professor,

Department of Mathematical Sciences,
Indian Institute of Technology (BHU) Varanasi,
Uttar Pradesh, INDIA 221005.

DECLARATION BY THE CANDIDATE

I, ***Manisha Singla***, certify that the work embodied in this Ph.D. thesis is my own bonafide work carried out by me under the supervision of ***Prof. K K Shukla*** and co-supervision of ***Dr. Debdas Ghosh*** from ***July 2017*** to ***March 2021*** at ***Department of Computer Science and Engineering***, Indian Institute of Technology (BHU) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, *etc.* reported in journals, books, magazines, reports, dissertations, theses, *etc.*, or available at websites and have not included them in this thesis and have not cited as my own work.

Date:

Place: Varanasi **(Manisha Singla)**

CERTIFICATE BY THE SUPERVISORS

It is certified that the above statement made by the candidate is correct to the best of our knowledge.

Supervisor

(Prof. K K Shukla)

Professor,

Department of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi.

Co-Supervisor

(Dr. Debdas Ghosh)

Assistant Professor,

Department of Mathematical Sciences,
Indian Institute of Technology (BHU) Varanasi.

Signature of Head of Department

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Robust SVM Variants and Their Applications

Name of the Student: Manisha Singla

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University) Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the *Doctor of Philosophy*.

Date:

Place: Varanasi **(Manisha Singla)**

Dedicated to my parents,

Mrs. Sunita Singla

and

Mr. Puran Chand Singla

ACKNOWLEDGEMENT

Though, only my name appears on the cover of this dissertation, so many great people have contributed to its production. I owe my gratitude to all those people who have made this thesis possible and because of whom my post graduate experience has been one that I will cherish forever.

I take this opportunity to express my profound gratitude and deep regards to my supervisor Dr. K. K. Shukla, Professor, Department of Computer Science and Engineering, IIT(BHU), Varanasi and co-supervisor Dr. Debdas Ghosh, Assistant Professor, Department of Mathematical Sciences, IIT(BHU), Varanasi for their exemplary guidance, monitoring and constant encouragement throughout the course of this dissertation. I am obliged to faculty members, in particular Prof. S. K. Singh, HoD, Department of Computer Science and Engineering, IIT(BHU), Varanasi and Prof. T. Som, HoD Department of Mathematical Sciences, IIT(BHU), Varanasi. In addition, I want to express my deepest gratitude to Mrs. Mridula Verma who supported me throughout my thesis work. I am grateful to Mr. Gaurav Bansal , who have provided me through technical, moral and emotional support during my research program.

A very special gratitude goes out to my colleagues and friends with special mention to Mrs. Anviti Pandey, Ms. Vandana Bharti, Ms. Pratishtha Verma, and Mr. Jitender Kumar.

Last but not the least, I specially thank my parents, Mr. Puran Chand Singla and Mrs. Sunita Singla, my parents-in-law Mr. Yogender Bansal and Mrs. Beena Bansal and my brother Ravi Singla for their constant support and encouragement, without which this assignment would have not been completed at all. I am grateful to my other family members who have supported me along the way.

(Manisha Singla)

Contents

List of Figures

List of Tables

List of Symbols

Symbol	Description
w	Weight Vector
b	Bias Term
λ	Regularization Parameter
e	Vector of ones
x	Input Vector
m	number of Instances
y	Target Vector
L_1	Absolute-value Norm
L_2	Euclidean Norm
L_K	Laplace Kernel Induced Loss Function
z	Margin Variable
$\hat{w}$	Optimal Weight Vector
$\hat{b}$	Optimal Bias Term
η	Rescaling Parameter of RSVM-PDProx
T	Total number of iterations
$\hat{\eta}$	Rescaling Parameter of Res-TSVR
τ	Parameter Used in Pinball Loss Function
α	Dual Variable
ϵ	Tolerance Value
R^2	Coefficient of Determination
$\hat{Y}$	Predicted Target Values
$\bar{Y}$	Mean of Y Values
L	Labeled Instances
U	Unlabeled Instances
ϕ	Kernel Function

Abbreviations

Abbreviation	**Description**
SVM	Support Vector Machine
PDProx	Primal-dual Proximal
TSVR	Twin Support Vector Regression
TWSVM	Twin Support Vector Machine
TSVM	Transductive Support Vector Machine
SGD	Stochastic Gradient Descent
ML	Machine Learning
SVR	Support Vector Regression
ECG	Electrocardiography
LS-SVM	Least-square SVM
FSVM	Fuzzy SVM
GEPSVM	Generalized Eigen Proximal SVM
PSVM	Proximal Support Vector Machine
OCSVM	One-class Support Vector Machine
CCCP	Concave-convex Procedure
DR	Diabetic Retinopathy
RSVR	Robust Support Vector Regression
DC	Difference of Convex
VAD	Voice Activity Detection
MRI	Magnetic Resonance Imaging
LRMD	Low-rank Matrix Decomposition
SV	Support Vectors
W-ETSVR	Weighted ϵ-TSVR
KKT	Karush-kuhn Tucker
NMSE	Normalized Mean Square Error
RMSE	Root Mean square Error
SSE	Sum Square Error

TSS	Total Sum Squares
SSR	Sum of Squares of Residuals
RBF	Radial Basis Function
SNR	Signal Noise Ratio
Ma	Microaneurysms
NPDR	Non-proliferative Diabetic Retinopathy
MPDR	Moderately Proliferative Diabetic Retinopathy
PDR	Proliferative Diabetic Retinopathy
SE	Soft Exudates
GMM	Gaussian Mixture Model
KNN	k- nearest Neighbor
ADHE	Adaptive Histogram Equalization
PDF	Probability Density Function
CNN	Convolutional Neural Network
PCA	Principal Component Analysis

PREFACE

Although support vector machines (SVMs) are one of the oldest machine learning approaches, both the regression and classification variants still help solve real-world tasks. However, the model has a limitation of sensitivity towards noise in the data set. The performance of SVM gets adversely affected when the model is trained with noisy data. This is due to the presence of unbounded hinge loss function. Since the other SVM variants also comprise hinge loss function, the limitation is inherited in all the SVM variants.

In this thesis, the robust formulations of SVM and its variants, which can handle sensitivity towards the noise, are proposed. This comes under the category of robust statistics. For classification tasks, the robustness against label noise in the data sets is added, while the robustness against Gaussian and uniform noise in the data sets is added for regression tasks.

First, the robustness against label noise in the conventional SVM is added. In this literature, the rescaled α-hinge loss function with a non-smooth L_2 regularizer isused. To solve the non-smooth optimization technique, the primal dual proximal (PDProx)-dual technique is implemented. The proposed approach is observed to be more sparse than the existing robust SVMs. The model converges at the rate of $O(1/T)$ where T denotes the number of iterations.

Next, the robustness against the uniform and Gaussian noise in a regression model, twin support vector regression (TSVR) has been added. The rescaled hinge loss func-

tion is used in this literature. To solve the non-convex problem, the half-quadratic optimization technique is used. Subsequently, an algorithm, Res-TSVR, has been devised to implement the proposed approach. It is found that the maximum number of iterations required to achieve an ϵ-precision solution is $O(\log(1/\epsilon))$.

Next, the twin support vector machine (TWSVM) is used in diabetic retinopathy detection using eye fundus images. In this work, the pinball loss function is used with TWSVM to add robustness to the model. It also helped in reducing the computational time. In all the experiments, the proposed approaches are compared with the existing approaches to prove the superiority of the proposed works.

The above contributions discussed are in the field of supervised machine learning. A semi-supervised machine learning framework has also been made robust to the label noise. Next contribution belongs to the addition of robustness to the conventional transductive support vector machine (TSVM). The semi-supervised learning model is robustified against label noise using a truncated pinball loss function. The model is tested on both small and large scale real-world data sets. The proposed approach is implemented using both the stochastic gradient descent (SGD) method and the dual problem solver, mlcv_quadprog (name is based on the machine learning-computer vision lab) . The robust model is also applied to the detection of COVID-19 infected patients using chest X-ray images.