Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2019) 27: 3247 – 3257
© TÜBİTAK
doi:10.3906/elk-1810-128

Importance-based signal detection and parameter estimation with applications to new particle search

Hatice DOĞAN[1*], **Nasuf SÖNMEZ**[2], **Güleser K. DEMİR**[1]
[1]Department of Electrical and Electronics Engineering, Faculty of Engineering, Dokuz Eylül University, İzmir, Turkey,
[2]Department of Physics, Faculty of Science, Ege University, İzmir, Turkey

Received: 18.10.2018 • **Accepted/Published Online:** 08.05.2019 • **Final Version:** 26.07.2019

Abstract: One of the hardest challenges in data analysis is perhaps the detection of rare anomalous data buried in a huge normal background. We study this problem by constructing a novel method, which is a combination of the Kullback–Leibler importance estimation procedure based anomaly detection algorithm and linear discriminant classifier. We choose to illustrate it with the example of charged Higgs boson (CHB) search in particle physics. Indeed, the Large Hadron Collider experiments at CERN ensure that CHB signal must be a tiny effect within the irreducible W-boson background. In simulations, different CHB events with different characteristics are produced and judiciously mixed with the non-CHB data, and the proposed method is applied. Our results show that distribution parameters of weak CHB signals can be estimated with high performance. This anomaly detection method is general enough to apply to similar problems in other fields (e.g., astrophysical, medical, engineering problems).

Key words: Anomaly detection, Kullback–Leibler importance estimation procedure, charged Higgs boson, importance estimation

1. Introduction

Standard model (SM) of fundamental particles is a theory that currently provides the best description of matter and its interactions at the Fermi scale [1]. There has been growing interest to get insights on "new physics" beyond the SM to discover "new particles" from the massive particle collision event data in recent decades. Generally, the main problem is to identify the extremely rare signals coming from the new particles from the huge background that originate from known physics. Several machine learning techniques such as neural networks, decision trees, and support vector machines have been used to identify these new particles in the literature [2–4]. These techniques pose the problem as a two-class classification problem. One of the classes is assigned for the events generated by the new particles, and the other one is assigned for the events coming from the SM. Parallel to the physics literature, we will call the first class the signal and the second one the background. Although two-class classification methods have success to a certain degree, they have various drawbacks regarding detecting events that are both rare and buried in huge and shadowing background data. The model dependency of the training data, unknown true model of the new particles, and ignorance of high similarities between signal and background events are some examples of these drawbacks.

In order to avoid these disadvantages, the problem of identifying the new particles is considered an

*Correspondence: hatice.dogan@deu.edu.tr

anomaly detection problem in this work. Anomaly detection is the problem of finding patterns in data that do not conform to an expected pattern in a dataset [5]. Anomaly detection can be applied in a wide variety of application domains such as fraud detection for credit cards [6, 7], intrusion detection for cyber-security [8], anomalous behavior detection for healthcare [9], and event detection for traffic scenes [10, 11], etc. In anomaly detection problems, generally there are considerably more normal data than the abnormal ones, and this situation fits perfectly with the detection of new particles in physics.

Due to the lack of prior knowledge on the anomaly data, most of the anomaly detection methods focus on the unsupervised problem setting where a labeled sample of training data is unavailable [5, 12–14]. These methods evaluate the samples in the low-density region as an anomaly. However, in the problem of searching for new particles, specifically, the particle charged Higgs boson (CHB) (in the two-Higgs-doublet model of type X [15]) is considered to be the signal (anomaly), and it is buried in a huge background (normal data). In other words, the anomalous data lies within the domain of background data, and it changes the data distribution collectively. Therefore, the unsupervised anomaly detection methods are not effective for the particle collision data. Since the distribution of the background is known in this problem, semisupervised methods are able to perform better than the unsupervised ones. Semisupervised methods generally assume that only normal data have labeling information [5, 16, 17]. These methods attempt to estimate the model of the normal data by using the training set and classify each data of the unlabeled test set as an anomaly if it seems unlikely to have been proposed by the process corresponding to the normal data distribution. In [16], a semisupervised anomaly detection method is proposed and applied to the signal classification problem in high energy physics. In this method, a mixture of Gaussians is fitted to a labeled background data, and then a mixture of this background model and a number of additional Gaussians are fitted to an unlabeled sample containing both background and signal. Fitting performance depends on the mixing ratio of the anomaly and the normal data, and it significantly reduces as the mixing ratio approaches the limits in realistic particle physics scenarios.

Since the new particles collectively change the distribution of the normal data, we pose the problem as finding the distribution differences between different data generation mechanisms, and then we directly estimate the ratio of two probability density functions, which is called “importance”, without estimating the densities. For this purpose, we use the Kullback–Leibler importance estimation procedure (KLIEP) [18–20] where importance is defined as the linear combinations of Gaussian kernels. In this study, we propose to use these importance values to detect the existence of CHB signal in a strong background. Additionally, we propose to use linear combination coefficients of importance for estimating the parameters of anomaly distribution to give crucial guidance to particle physicist in setting up the search process. To our knowledge, no previous study has been presented for using these coefficients for classification or parameter estimation purposes. We consider the kinematical properties of the particles which are produced as a result of the collision. The invariant mass distribution of the W bosons forming the SM background is chosen as the normal distribution, and the SM background that has the possibility to contain CHB signal is chosen as the distribution to be tested. Since the importance estimated by KLIEP deviates from the one value according to the mean and the variance of the invariant mass distribution of CHB buried in the background, a useful and effective approach to find the unknown parameters of the signal distribution has been proposed.

The rest of the paper is organized as follows. In Section 2, the importance-based anomaly detection method is presented. In Section 3, detection of the charged Higgs signal and prediction of its characteristics are discussed, and the results are presented. Finally, the discussions are delivered in Section 4.

2. Importance-based anomaly detection

In this section, we propose to use the importance values of two density distributions for determining an anomaly in the particle collision events. If there is an anomaly, clearly, it indicates the existence of a new particle. Apart from the conventional classification approach where the decision boundaries between the classes are determined for the discriminative features, we aim to find the anomalies by examining the density distributions. Since the density distribution of the SM background is known, the deviation from this distribution is considered an anomaly. Here it should be noted that the anomalies are rare; therefore, the test distributions containing anomalies do not significantly differ from the SM distribution. In the KLIEP approach, anomalies are found based on the training set consisting only of the background data. The KLIEP algorithm does not model the training and the test set densities directly, but only estimates the ratio of the training to the test data densities, which is called importance [18–20].

Let, $D \in R^d$ be the data domain, the training samples $\{x_j^{tr}\}_{j=1}^{n_{tr}}$ from a training distribution with density $p_{tr}(x)$ and test samples $\{x_i^{te}\}_{i=1}^{n_{te}}$ from a test distribution with density $p_{te}(x)$ are given. The training dataset consists of samples from normal data while test set contains some anomalous data. The importance, $w(x)$ can be defined as

$$w(x) = \frac{p_{tr}(x)}{p_{te}(x)} \tag{1}$$

and the linear model of importance, $\hat{w}(x)$ can be given by

$$\hat{w}(x) = \sum_{l=1}^{b} \alpha_l \phi_l(x), \tag{2}$$

where $\{\alpha_l\}_{l=1}^{b}$ are the parameters to be learned from the data samples, and $\{\phi_l(x)\}_{l=1}^{b}$ are basis functions such that $\phi_l(x) \geq 0 \;\; \forall x \in D$ and $l = 1, 2, ..., b$. Using the model $\hat{w}(x)$, an estimator of training density, $\hat{p}_{tr}(x)$ can be written as

$$\hat{p}_{tr}(x) = \hat{w}(x) \; p_{te}(x). \tag{3}$$

The $\{\alpha_l\}_{l=1}^{b}$ parameters can be determined by minimizing the Kullback–Leibler divergence from $p_{tr}(x)$ to $\hat{p}_{tr}(x)$:

$$\begin{aligned} KL[p_{tr}(x)||\hat{p}_{tr}(x)] &= \int_D p_{tr}(x) \log \frac{p_{tr}(x)}{p_{te}(x) \; \hat{w}(x)} \; dx \\ &= \int_D p_{tr}(x) \log \frac{p_{tr}(x)}{p_{te}(x)} \; dx - \int_D p_{tr}(x) \log \; \hat{w}(x) \; dx. \end{aligned} \tag{4}$$

Since only the second term in Eq. 4 depends on $\{\alpha_l\}_{l=1}^{b}$, the function that is optimized can be rewritten as

$$KL' = \int_D p_{tr}(x) \log \; \hat{w}(x) \; dx \approx \frac{1}{n_{tr}} \sum_{j=1}^{n_{tr}} log \; \hat{w}(x_j^{tr}). \tag{5}$$

This is the objective function to be maximized with respect to parameters $\{\alpha_l\}_{l=1}^{b}$ [20]. Since $w(x)$ is nonnegative by definition, the estimation $\hat{w}(x) \geq 0$ for all $x \in D$, which can be achieved by restricting $\alpha_l \geq 0$

for $l = 1, 2, \cdots, b$. $\hat{w}(x)$ should also be normalized because $\hat{p}_{tr}(x)(= \hat{w}(x)p_{te}(x))$ is a probability density function:

$$1 = \int_D \hat{p}_{tr}(x)\ dx = \int_D \hat{w}(x)p_{te}(x)\ \ dx \approx \frac{1}{n_{te}} \sum_{i=1}^{n_{te}} \hat{w}(x_i^{te}) \tag{6}$$

Now the convex KLIEP optimization criterion can be given as follows:

$$\begin{aligned} &\max_{\{\alpha_l\}_{l=1}^{b}} \left[\sum_{j=1}^{n_{tr}} \log \left(\sum_{l=1}^{b} \alpha_l \Phi_l \left(x_j^{tr} \right) \right) \right] \\ &s.t.\ \frac{1}{n_{te}} \sum_{i=1}^{n_{te}} \sum_{l=1}^{b}\ \alpha_l\ \Phi_l \left(x_i^{te} \right) = 1\ and\ \alpha_1,\ \alpha_2,\ ...,\ \alpha_b \geq 0. \end{aligned} \tag{7}$$

The global solution can be obtained by simply performing gradient ascent and feasibility satisfaction iteratively [20]. The Gaussian Kernel model centered at the training input points x_j^{tr} can be chosen as simple yet effective basis functions $\{\Phi_l(x)\}_{l=1}^{b}$, and kernel width can be optimized with likelihood cross validation procedure.

If the $p_{te}(x)$ distribution does not contain anomalous data, the estimated importance values $\hat{w}(x)$ will be close to one; otherwise, these values will deviate from one. We illustrate how the KLIEP algorithm behaves in anomaly detection in Figure 1, and the training density is taken to be $p_{tr}(x) = N(x; 0, 1)$ as the normal distribution where $N(x; \mu, \sigma)$ denotes the Gaussian density with mean μ and variance σ^2. We choose $n_{tr} = 1000$ training samples and $n_{te} = 1000$ test samples from $p_{tr}(x)$ randomly, and we add 50 anomalous data to the test set. The number of basis functions in the KLIEP is fixed to $b = 100$. The histograms of the datasets, as well as the estimated importance values obtained by KLIEP, are depicted in Figure 1. The graph shows that the importance values $\hat{w}(x)$ are small in the anomalous data range, and therefore, the anomaly can easily be detected.

Since the anomalous data collectively change $p_{te}(x)$ in our problem, the importance values can be discriminative for the identification of different anomaly groups. To illustrate this point, we experiment with United States Postal Service Handwritten Digit Database (USPS) that consists of 16x16 images of handwritten digits taken from US mail envelopes [21]. In the experimental setup, we take 1200 images of ‘1’ as the training set (background) and consider the images of ‘0’, ‘3’, ‘7’, and ‘9’ as the different anomaly groups. Figure 2 shows some samples of the digits used. There are 30 test sets for each anomaly digits obtained by mixing 1200 images of ‘1’s and 60 randomly selected images of an anomaly digit. The importance values are calculated for each dataset, and the mean of the importance values for each digit is computed. Minimum and maximum values of the mean values of the background and the anomaly data are also calculated to show the discriminative capabilities of importance values. The results are given in Table. In the table, the labels A and B denote the importance values of the anomaly and the background data, respectively. The digits ‘7’ and ‘9’ are similar to the digit ‘1’ so their importance values are higher than the importance values of ‘0’ and ‘3’. When the test set contains only the digit ‘1’, the importance values get close to one.

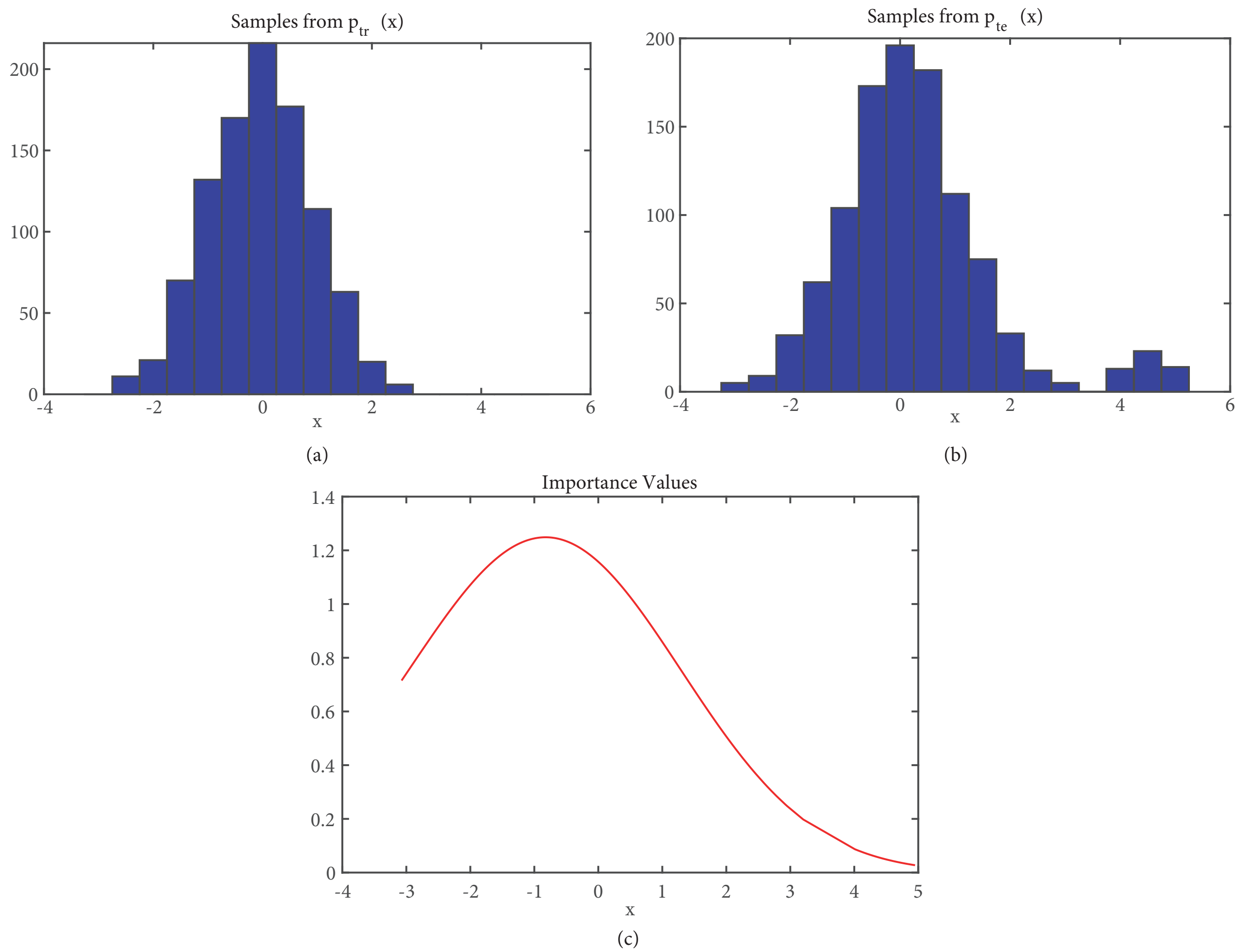

Figure 1. The distribution of samples drawn from (a) $p_{tr}(x)$, (b) $p_{te}(x)$, (c) estimated importance values.

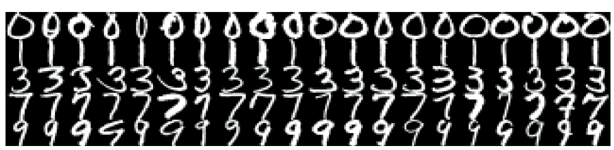

Figure 2. USPS data samples.

Table. Minimum, mean, and maximum values of importance for different datasets.

	B_{min}	B_{mean}	B_{max}	A_{min}	A_{mean}	A_{max}
Zero	0.672	1.047	1.196	0.012	0.065	0.271
One	0.772	1.000	1.087	0.819	0.999	1.076
Three	0.612	1.045	1.213	0.017	0.096	0.320
Seven	0.548	1.044	1.227	0.025	0.125	0.418
Nine	0.528	1.043	1.228	0.018	0.151	0.487

3. Search for the charged Higgs boson

3.1. Data generation

Although there are various theoretical schemes, two-Higgs-doublet model (2HDM) [22], which can be considered as a minimal extension of the SM, is used in this work to illustrate the anomaly detection capability of the proposed approach. 2HDMs have been studied for more than a couple of decades extensively due to their salient features [15, 23]. Our aim here is to study the charged Higgs signal in top quark decays [15, 23] produced in proton–proton collisions. We primarily focus on the single-top quark production via ($pp \to X\ t \to X\ b\ H^{\pm}$) channel. Top-quark could decay through the W and the charged Higgs (H). Therefore, decaying via W-boson of the SM generates the irreducible SM background.

Free parameters of the 2HDM are specified, and necessary model files for event generation are calculated with the help of 2HDMC software [24]. Event generation and simulation are obtained using MADGRAPH software, which simulates proton–proton collisions using the model files [25].

The mean of the invariant mass distribution of the SM background, μ_W, is 80 GeV. As noted before, the mean of the invariant mass distribution of the CHB signal, μ_H, is unknown. However, considering the literature on particle physics it is reasonable to expect μ_H between 90 and 120 GeV [15, 26]. Therefore, in this study, to generate anomaly data, four different μ_H values are chosen; specifically $\mu_H = \{90, 100, 110, 120\}$ GeV. The CHB signal is generated for each μ_H value considering two different variance cases. The first case is where the variance of the CHB signal is smaller than the variance of the SM background ($\sigma_H < \sigma_W$), and the other one is equal to it ($\sigma_H = \sigma_W$).

The MADGRAPH event generator program is run for each μ_H and σ_H combination. Therefore, eight different CHB datasets have been produced. The datasets are formed by mixing the CHB dataset with 100,000 randomly selected data from the SM background using the mixing ratios of $1\%, 3\%$, and 5%. 100 different mixed datasets are obtained for each μ_H and σ_H values for test purposes.

3.2. Higgs signal detection

For the detection of the CHB signal, we calculate the importance values for the datasets that contain the signal with different characteristics and ratios by using the KLIEP algorithm. The background datasets are chosen as the samples of $p_{tr}(x)$ and contaminated sets are taken as the samples of $p_{te}(x)$. The centers of the Gaussian kernels are chosen in the range of background data with equal intervals, and the optimal number of the kernel, b, is found as 100, as a result of extensive simulation studies. Model selection in KLIEP is done by likelihood cross validation, as suggested in [20].

As explained in Section 2, the importance values estimated by the KLIEP must be close to one if the test set does not contain an anomaly. To check this, we choose training samples and test samples from the background randomly, and we add samples again from the background to the test set with %1 ratio. The importance value versus invariant mass (m) graph is shown in Figure 3. The results show that if the data does not contain an anomaly, all importance values are close to one as expected.

In order to observe the effect of the mixture ratio of the CHB signal on the importance values, the KLIEP is run for the contaminated background sets that have the CHB signal with 1%, 3%, and 5% mixing ratios, and the results for $\mu_H = 90$ GeV are shown in Figure 4. It can be seen from the figure, as the contamination ratio increases, the estimation of importance values decay rapidly. It is expected since as the ratio of anomaly increases, the test distribution will be more different than the normal distribution.

Additionally, the importance values seem to increase in the region where the anomaly is absent or rare because of the attenuated test distribution. In the following analysis, we focus on the performance of the proposed approach with 1% mixing ratio to present a more applicable method for the experimental data in particle physics.

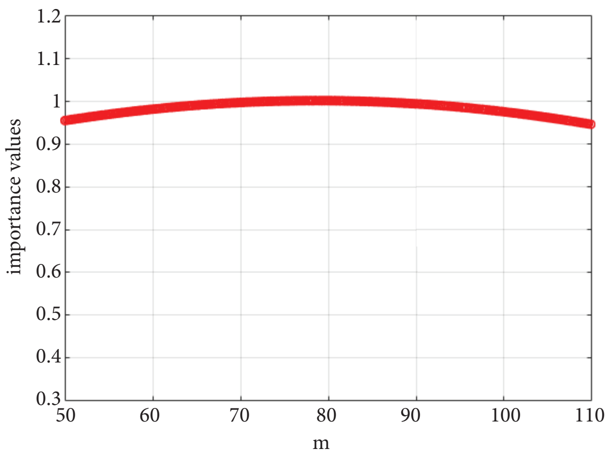

Figure 3. The importance values between random background datasets.

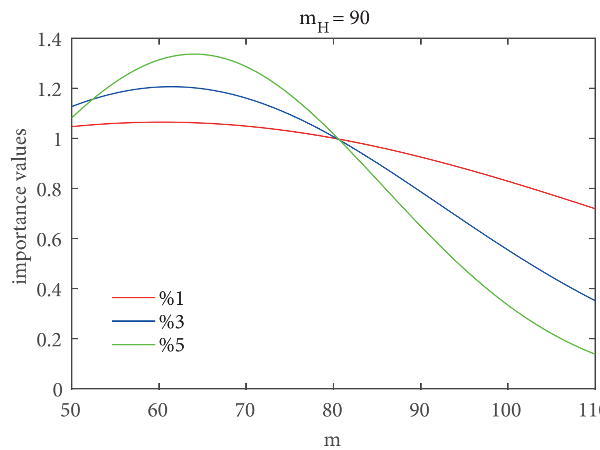

Figure 4. The importance values of contaminated background sets that have the Higgs signal with 1%, 3% ve 5% ratios.

Figures 5a and 5b show the importance values for the datasets that contain the CHB signal distributions with different μ_H values for the following cases: $\sigma_H < \sigma_W$ and $\sigma_H = \sigma_W$. The least deviation from the value of one has been observed for $m_H = 90$ GeV. This is due to the fact that the signal samples are clustered in the region where the background data is very dense. It can be seen from the graphs that the importance decays faster when the signal samples are clustered in the regions where the background data is less dense as in the cases $\mu_H = 110$ and $\mu_H = 120$ GeV. The effect of the variance difference on the importance values can be seen much clearer in Figure 6 for $\mu_H = 90$ GeV. The importance values of equal variance case ($\sigma_H = \sigma_W$) are lower than the narrow variance case ($\sigma_H < \sigma_W$) at the anomaly region. Here, the wider spread of the signal data might have a higher effect on the less dense area of normal distribution. The same statement is also valid for the other charged Higgs contaminated background sets.

3.3. Estimation of distribution parameters

The importance values estimated by the KLIEP method deviate from the value of one when the test set contains the CHB signal. Additionally, it is observed that this deviation depends on the mean μ_H and variance σ_H values of the CHB distribution. According to this observation, the characteristics of the importance functions will be similar if the test sets contain the CHB data that have the same μ_H and σ_H values. If this characteristic can be determined, the mean and the variance of the 2HDM data contained in the SM model can be found by the amount of deviation from the importance function.

From Eq. 2, it can be seen that the importance values are the linear combination of Gaussian kernels, and α parameters will form the importance. Therefore, it is proposed to use nonzero α values as the features for the classification of different CHB anomaly characteristics. We have 100 sets for every contamination, while half of them is used for training the other half is used for testing.

Different classifiers such as K-nearest neighbor (KNN), support vector machines (SVM), and linear

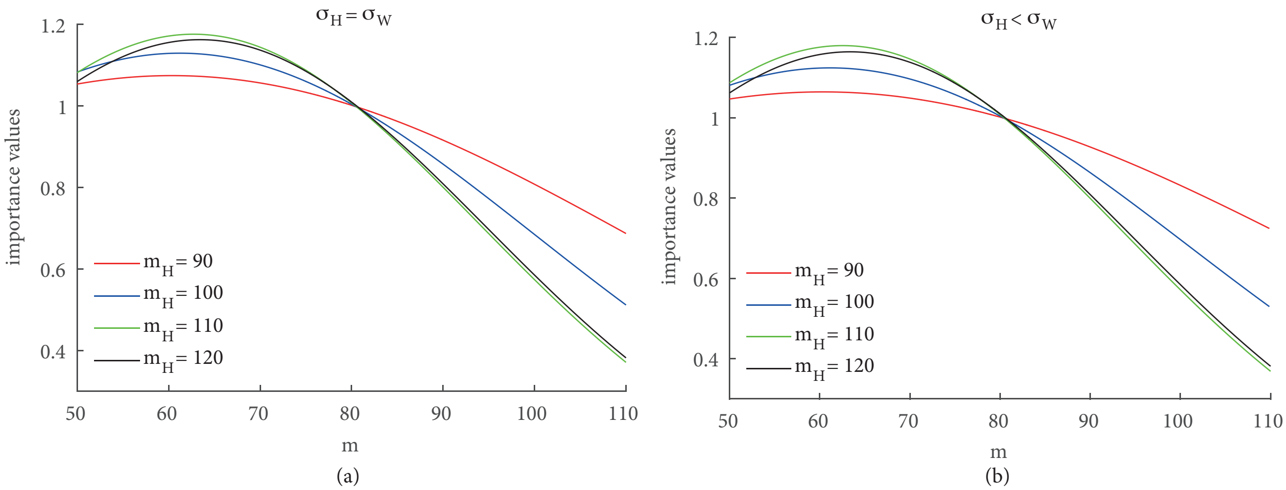

Figure 5. The importance values of contaminated background sets that have the CHB signal distribution with $\mu_H = \{90, 100, 110, 120\}$ values.

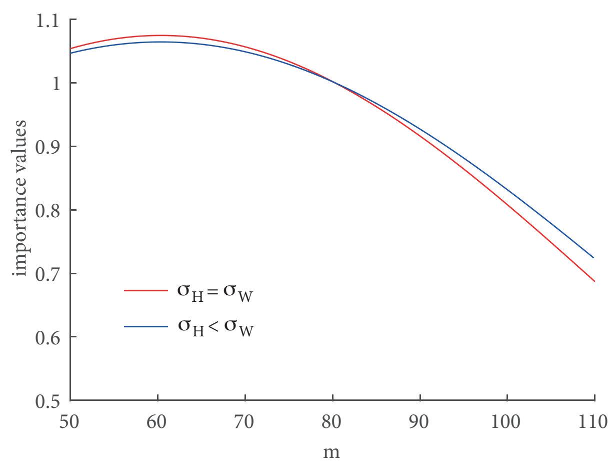

Figure 6. The importance values of contaminated background sets that have the CHB signal distribution with same mean but different variance values.

discriminant (LD) classifiers [27] are employed for the estimation of anomaly distribution parameters. The classifiers are trained by using 5-fold cross validation with optimized parameters. Obtained confusion matrices and accuracies of classifiers are given in Figure 7. In the figure, T and O denote the target and output classes, respectively, and while the superscript indicates μ_H values, subscript indicates variance states. For example, T_1^{90} stands for the target classes of Higgs distribution that has $\mu_H = 90$ GeV, and the variance $\sigma_H = \sigma_W$ and O_2^{90} stands for the output classes of Higgs distribution that has $\mu_H = 90$ GeV and the variance $\sigma_H < \sigma_W$. Clearly, there are eight different classes, each corresponding to the different parameter pairs (μ_H, σ_H) of anomaly distribution.

It can be seen from the results that all the classifiers classify μ values of the charged Higgs distributions correctly, which is buried in the huge background, by using the α parameters of the importance curves. The performance of the KNN and the SVM classifiers are not as good as the LD classifier in detecting the variance differences of the distributions that have the mean values of $\mu_H = 110$ GeV and $\mu_H = 120$ GeV. LD classifier

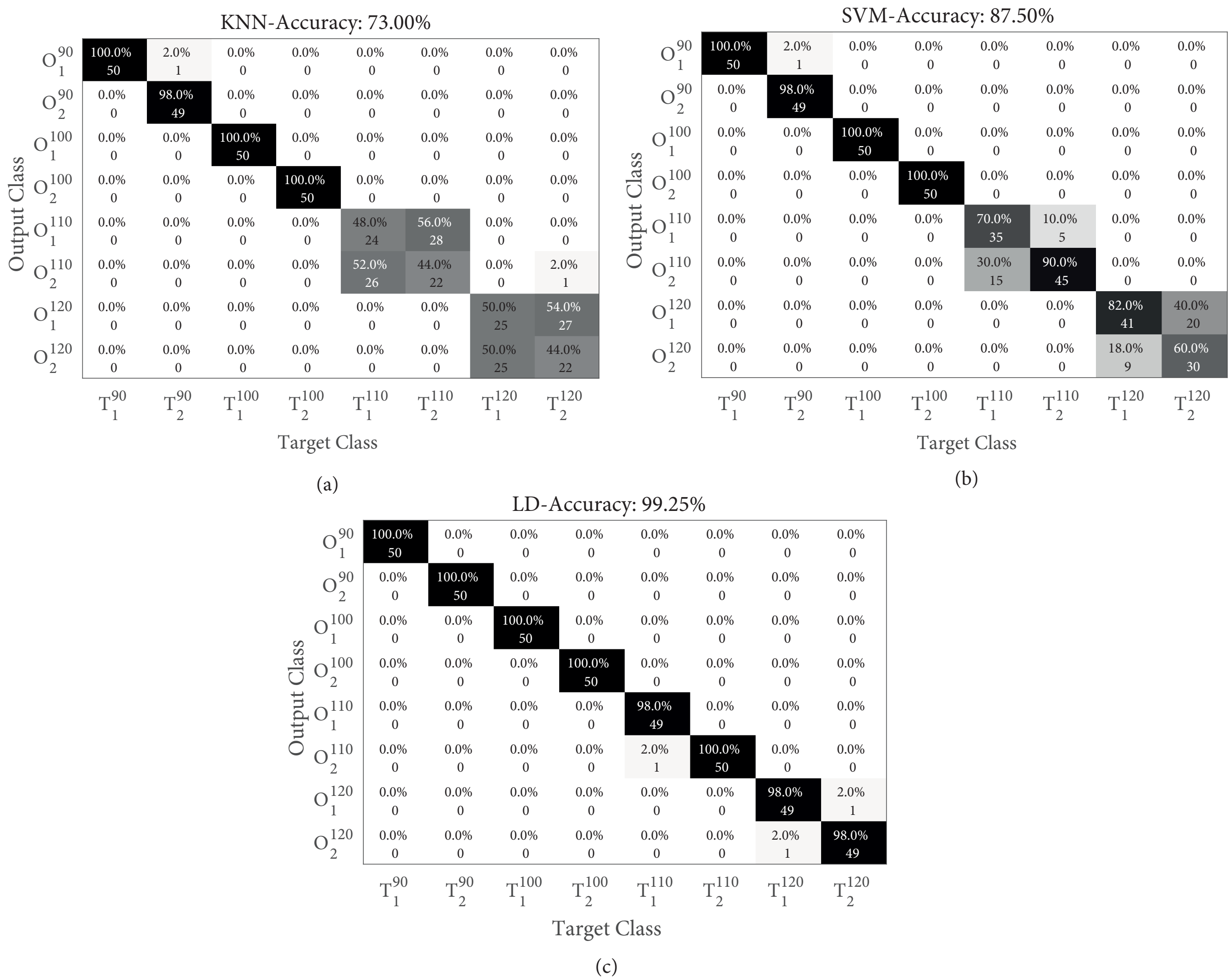

Figure 7. Confusion matrices of (a) KNN, (b) SVM, and (c) LD classifiers.

correctly classifies the importance curves with 99.25% accuracy. This confirms the result that the information about the characteristics of the charged Higgs signal distribution buried in the SM background signal can be extracted from the importance values.

Since we are focusing on not point anomaly, but a collective one, and we are targeting to estimate the anomaly distribution parameters, not to classify each anomaly data, the problem at hand has a distinct nature. We find only the study by Vatanen [16] as the closest one to our approach. Although [16] used semisupervised collective anomaly detection method considering the distribution differences, the main purpose was to classify the anomalies. The authors did not give any results for finding the parameters of the distribution. Additionally, they stated that the robustness of their classifying performance starts to suffer when the test data contains less than 3% of the signal.

4. Discussion

Detection of a charged Higgs particle, if any, is one of the main concerns of the Large Hadron Collider in RUN II phase. Currently, there is no signal from new particles like the CHB, and assessment is that their interactions must be too weak to show a head above the background. For the purpose of enhancing the extraction performance, we focus on a more fundamental question: How to know the existence and the possible regions of rare signals in strong backgrounds? As shown and experimented in the body of the paper, we have been able to answer this question by the importance-based detection and the estimation method that we propose. We predicted the range of the signal distribution which gives crucial guidance to particle physicist in setting up to search process.

In this study, the proposed method is employed to solve a challenging problem where anomalies occur as a cluster in the background domain. Indeed, this method can also be applied to those where anomaly data is not necessarily in the dense background domain. Disease detection from heart rate signals, novelty detection in large-scale astronomical datasets, and estimation of fault events for predictive maintenance are a few examples of possible application areas.

Acknowledgments

This work is supported by the TÜBİTAK grant 113F167.

References

[1] Tanabashi M, Hagiwara K, Hikasa K, Nakamura K, Sumino Y et al. Review of particle physics. Physical Review D 2018; D98 (1): 030001. doi: 10.1103/PhysRevD.98.030001

[2] Bhat PC. Multivariate analysis methods in particle physics. Annual Review of Nuclear and Particle Science 2011; 61: 281-309. doi: 10.1146/annurev.nucl.012809.104427

[3] Carli T, Koblitz B. A multi-variate discrimination technique based on range-searching. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 2003; 501 (2-3): 576-588. doi: 10.1016/S0168-9002(03)00376-0

[4] Wolter M. Multivariate analysis methods in physics. Physics of Particles and Nuclei 2007; 38 (2): 255-268. doi: 10.1134/S1063779607020050

[5] Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Computing Surveys (CSUR) 2009; 41 (3): 15:1-15:58. doi: 10.1145/1541880.1541882

[6] Aleskerov E, Freisleben B, Rao B. Cardwatch: a neural network based database mining system for credit card fraud detection. In: IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr); New York City, NY, USA; 1997. pp. 220-226.

[7] Mohiuddin A, Mahmood AN, Islam MR. A survey of anomaly detection techniques in financial domain. Future Generation Computer Systems 2016; 55 (C): 278-288. doi: 10.1016/j.future.2015.01.001

[8] Kumar V. Parallel and distributed computing for cybersecurity. IEEE Distributed Systems Online 2005; 6 (10): 1-9. doi: 10.1109/MDSO.2005.53

[9] Liao Z, Kong L, Wang X, Zhao Y, Zhou F et al. A visual analytics approach for detecting and understanding anomalous resident behaviors in smart healthcare. Applied Sciences 2017; 7 (3): 254. doi: 10.3390/app7030254

[10] Riveiro M, Lebram M, Elmer M. Anomaly detection for road traffic: a visual analytics framework. IEEE Transactions on Intelligent Transportation Systems 2017; 18 (8): 2260-2270. doi: 10.1109/TITS.2017.2675710

[11] Yuan Y, Wang D, Wang Q. Anomaly detection in traffic scenes via spatial-aware motion reconstruction. IEEE Transactions on Intelligent Transportation Systems 2017; 18 (5): 1198-1209. doi: 10.1109/TITS.2016.2601655

[12] Breunig MM, Kriegel HP, Ng RT, Sander J. Lof: identifying density-based local outliers. In: ACM SIGMOD International Conference on Management of Data; Dallas, TX, USA; 2000. pp. 93-104. doi: 10.1145/335191.335388

[13] Tang J, Chen Z, Fu AWC, Cheung DW. Enhancing effectiveness of outlier detections for low density patterns. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining; Taipei, Taiwan; 2002. pp. 535-548. doi: 10.1007/3-540-47887-6_53.

[14] Schölkopf B, Williamson RC, Smola AJ, Shawe-Taylor J, Platt JC. Support vector method for novelty detection. In: Advances in Neural Information Processing Systems 12; Denver, CO, USA; 1999. pp. 582-588.

[15] Guedes R, Moretti S, Santos R. Charged higgs bosons in single top production at the LHC. Journal of High Energy Physics 2012; 10: 119. doi: 10.1007/JHEP10(2012)119

[16] Vatanen T, Kuusela M, Malmi E, Raiko T, Aaltonen T et al. Semi-supervised detection of collective anomalies with an application in high energy particle physics. In: The 2012 International Joint Conference on Neural Networks (IJCNN); Brisbane, QLD, Australia; 2012. pp. 1-8.

[17] Aissa NB, Guerroumi M. Semi-supervised statistical approach for network anomaly detection. Procedia Computer Science 2016; 83: 1090-1095. doi: 10.1016/j.procs.2016.04.228

[18] Sugiyama M, Nakajima S, Kashima H, Buenau PV, Kawanabe M. Direct importance estimation with model selection and its application to covariate shift adaptation. In: Advances in Neural Information Processing Systems 20; Vancouver, BC, Canada; 2007. pp. 1-8.

[19] Sugiyama M, Suzuki T, Nakajima S, Kashima H, Buenau PV et al. Direct importance estimation for covariate shift adaptation. Annals of The Institute of Statistical Mathematics 2008; 60 (4): 699-746. doi: 10.1007/s10463-008-0197-x

[20] Hido S, Tsuboi Y, Hashima H, Sugiyama M, Kanamori T. Statistical outlier detection using direct density ratio estimation. Knowledge and Information Systems 2011; 26 (2): 309-336. doi: 10.1007/s10115-010-0283-2

[21] Hull JJ. A database for handwritten text recognition research. IEEE Transactions on Pattern Analysis and Machine Intelligence 1994; 16 (5): 550-554. doi:10.1109/34.291440

[22] Branco GC, Ferreira PM, Lavoura L, Rebelo MN, Sher M et al. Theory and phenomenology of two-higgs-doublet models. Physics Reports 2012; 516 (1-2): 1-102. doi: 10.1016/j.physrep.2012.02.002

[23] Hashemi M. Observability of light charged higgs decay to muon in top quark pair events at LHC. European Physical Journal 2012; 72 (5): 1994. doi: 10.1140/epjc/s10052-012-1994-1

[24] Eriksson D, Rathsman J, Stal O. 2HDMC: Two-higgs-doublet model calculator. Computer Physics Communication 2010; 181 (1): 189-205. doi: 10.1016/j.cpc.2009.09.011

[25] Alwall J, Artoisenet P, de Visscher S, Duhr C, Frederix M et al. New developments in MadGraph/MadEvent. In: 16th International Conference on Supersymmetry and the Unification of Fundamental Interactions; Seoul, Korea; 2008. pp. 84-89. doi: 10.1063/1.3052056

[26] Moretti S. 2HDM charged higgs boson searches at the LHC: status and prospects. In: Prospects for Charged Higgs Discovery at Colliders; Uppsala, Sweden; 2016. doi: 10.22323/1.286.0014

[27] Webb AR, Copsey KD. Statistical Pattern Recognition. 3rd ed. West Sussex, UK: John Wiley & Sons, 2011.