

Mother's Lifestyle Feature Relevance for NICU and Preterm Birth Prediction

Himani Deshpande^{1,*}, Leena Ragha^{2**}

^{1,2} Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, India

Abstract. Maternal health plays an important role in defining the health of mother, child and childbirth experience. With the change in lifestyle over the decades, there have been many health challenges faced by woman, which makes it important for women to understand the impact of their lifestyle and physical health features on their wellbeing. In this study, we have realised the importance of mother's features with respect to preterm childbirth prediction and prediction for neonatal intensive care unit(NICU) facility requirement for newborn. Experiments are performed on MSF dataset which consists of records of 1000 women, 21 physical features and 78 lifestyle features are taken into consideration. Random forest based hybrid model using F-score and Mutual information is used to evaluate each features for their capability of True positive(TP) and False Negative(FN) predictions. For preterm birth prediction, out of all the features hypertension, diabetes, PCOS and consumption of outside food during teenage are found to be the most relevant features. While for NICU prediction diabetes, low amniotic fluid during pregnancy, exposure to air and noise pollution during teenage and consumption of alcohol after marriage are found to be relevant.

1 Introduction

Women are considered to be the key reason behind a happy family and a progressive nation. They are appreciated for carrying out multiple roles with great efficiency, one such responsibility is childbirth. Women are said to be the creators, and have been awarded with immense strength and ability to reproduce. She needs to be strong physically as well mentally to carry the baby in her womb and to enable a healthy pregnancy and childbirth experience. Her health has always been considered as one of the main goal of countries worldwide, as a healthy mother gives birth to a healthy child.

Woman's health is getting effected with rapid urbanisation as there are many challenges that women are facing which includes addition responsibilities and increased societal expectations[1]. Woman is supporting her family emotionally, physically and financially. She is breaking old stereotypes and boundaries and have proved her worth across all the domains. With all these comes a new set of challenges which results in change in her daily routine, stress levels, social environment and physical health status. A woman's health is closely associated with the way she lives through different phases of her life, which is called as her lifestyle. There have been changes in terms of her dietary pattern, sleeping patterns, travelling routine, job profiles, ignorance towards her health, delaying of pregnancy, dependence upon medicines, consumption of unhealthy

substances like caffeine, alcohol, nicotine, etc. All these things are slowly effecting her health [2-5], according to a survey carried out by the Associated Chamber of Commerce and Industry (ASSOC-HAM), around 68% of women in the age group 21-52 years, who are working are afflicted with lifestyle related health issues like anxiety, diabetes, depression, reproductive issues, hypertension and obesity[6]. Woman's fertility and childbearing experience is influenced by her lifestyle[7,8] across her reproductive age which starts from the onset of menses during adolescent years. It is believed that infant's health is effected by the way a mother lives through different phases of her life specially during the time of pregnancy. Woman are advised to be careful and follow healthy habits during pregnancy. Thus making it important for her to understand her lifestyle and physical features which could have an effect on her health, child's health and overall childbirth experience. There are many things which together sum up her daily routine, she needs to realise till what extent these factors influence her reproductive health. For a healthier pregnancy along with preconception health care it is advised to understand the interconnection between various influencing factors [9]. Health complications can be minimised scientifically by machine learning computational models by finding the impact of features on woman and child health.

Machine learning has emerged as a scientific tool for understanding data and identifying the association of features with outcomes. Feature selection methods have

* Corresponding author: himani.deshpande@thadomal.org ** leena.ragha@rait.ac.in

gained popularity as machine learning models work accurately when informative data is provided as input while redundant and irrelevant data is filtered out. Feature selection helps to understand the importance of features while finding solutions. Feature selection methods are broadly divided into three categories namely filter method, wrapper method and embedded method. Woman's lifestyle and physical features could be evaluated using these methods, thus helping her to know where she needs to be careful in order to avoid future health complications.

The need of the hour is to create awareness among the female population of the society regarding their reproductive health as well as overall wellbeing and to educate them about the factors that affect their general and emotional wellbeing, and specifically the reproductive health. The aim is to prevent pregnancy complications to ensure a happy and meaningful life to mother and child.

2 Literature Review

Features plays an important role in medical domain, making feature selection a popular practice with medical data. For feature selection there is no significant work in maternal domain thus we have looked into research work across other domains. Yi- Wei et. al.[10], selected best few features using F-score, selected features were given as input to Random forest algorithm to further filter out relevant features. Paul et.al.[11] proposed GARF (Genetic Algorithm based on Random forest) for ranking features available in positron emission tomography clinical data and images. Features were evaluated initially using spearman's correlation following by Genetic algorithm. Population selected by genetic algorithm is further weighted using Random forest misclassification rate, AUC and sparsity constraints. Peng et. al.[12], used minimum redundancy maximal relevance criterion for evaluating features, followed by backward and forward wrapper selection using multiple classification algorithms. Lee et.al.[13], first used genetic algorithm with dynamic parameter (GADP) settings for microarray data analysis. GADP differs from the traditional method in terms of setting the mutation and crossover rates. Huang et.al.[14] used a hybrid model, where two feature sets are generated using F-score and Information gain separately, these sets were then combined and given as input to SVM based wrapper method. P. Jaganathan et.al.[15], used modified F-score for more than two classes. The threshold is set to be mean of Improved F-score calculated for each feature. Huijuan et.al.[16], proposed MIMAGA-Selection algorithm for gene expression data. Harun et. al.[17],

worked on text data using two-stage process, Information gain is used to evaluate text, most important features found are given as input to PCA and genetic algorithm for evaluation. Zheng et. al.[18], selected best features from the Diabetes Mellitus dataset using hybrid approach.

3 Methodology

This work focus upon understanding importance of mothers lifestyle and physical features in predicting the childbirth outcome being preterm or full-term and predicting if there will be a need to provide neonatal intensive care unit (NICU) for the newborn. In this section we will be looking into the methodology used to achieve this aim.

3.1 Dataset

Through the literature survey we realised that there are no standard datasets available and hence we created our own dataset named Mother's Significant features(MSF) [19] dataset which we used in this study. MSF dataset consists of 1000 records of mothers, this dataset comprise of detailed information on mother's lifestyle during her teenage, after marriage and during her pregnancy. Women residing in Mumbai metropolitan region of India, were interviewed by medical personnel just after childbirth. A total of 21 physical features and 78 lifestyle features were taken into consideration. Lifestyle features includes information about their home and work routine, economical status and stress related information. Out of the available records, 172 childbirth were born premature and 828 were full term deliveries. As per MSF dataset records, 243 newborn were availed NICU facility while 757 didn't avail this facility.

3.2 Pre-processing

MSF dataset consists of a mix of continuous and categorial features. Decision tree is biased towards continious features and features with more categories [20], for the same reason all the continuous features were converted into categorial with similar number of categories. These categories were decided after looking into literature and consulting gynaecologists and paediatricians.

3.3 VIBRF

The algorithm Variation and Information based random forest (VIBRF) is proposed by us and explained in detail in [21] is used for lifestyle relevant feature weighing in the proposed study. VIBRF is a hybrid feature selection model which combines filter method and embedded

method to evaluate each feature. F-score is used to find the variation across the feature while entropy and mutual information estimates the information provided by the feature. Each feature is ranked looking into both the factors simultaneously. Random forest is formed using hybrid DF_trees, out of all these trees, most accurate were considered. Root nodes of these DF_trees were considered to be most relevant feature of that tree. Available features were found to be relevant if they appear to be root node of these trees, each features is assigned weights looking into the number of DF_Trees it is assigned as root node for.

3.4 Feature Importance Weighing

With this work we have assigned the weights to mother’s features. Features are associated with their capability to predict occurrence of birth complications. Mother’s physical health at the time of pregnancy is analysed using physical feature set of MSF dataset. Features defining mother’s way of living, her family and environment conditions are covered under lifestyle sub-dataset of MSF dataset. Lifestyle features used for the study are spread over threes phases of her reproductive age that are firstly during her teenage, secondly after marriage and thirdly during pregnancy.

In the proposed model we have created random forests using 1000 DF_Trees, nodes of the tree is selected using the equation (1), where x and y are taken as 40% and 60% respectively, giving more weitage to mutual information achieved by the features while predicting outcome. FG_N is the Final Gian for Nth feature , F-Score $_N$ is the F-score for Nth feature and MI_N is entropy based mutual information. Equation (2), defines the entropy for Nth features used in equation (1).

$$FG_N = (x\%) F-Score_N + (y\%)MI_N \quad (1)$$

$$Entropy(N) = -p(yes) \log_2 p(yes) - p(no) \log_2 p(no) \quad (2)$$

In this study features are weighted on their capability of predicting true positive and false negative. VIBRF model is implemented twice, firstly random forest (VIBRF_TP) is formed using DF_Tree, best tress are selected having highest True Positive (TP) prediction and secondly VIBRF_FN consisting of best DF_Trees which are ones having highest False negative(FN) values. All features are given weights for being relevant considering root nodes of VIBRF_TP random forest and for being the misleading features looking into root nodes of VIBRF_FN. In the next section we will be discussing the weight given to each feature.

4 Results

Experiments are performed on mother’s physical features and lifestyle features to find feature relevance for predicting preterm birth and NICU requirement for newborn. Features weightage using VIBRF_TP predictions were considered to be the relevance weight of features, while features weight using VIBRF_FN random forest is misclassification weight of feature. Fig. 1, shows the relevance and misclassification weight of physical features while predicting preterm childbirth.

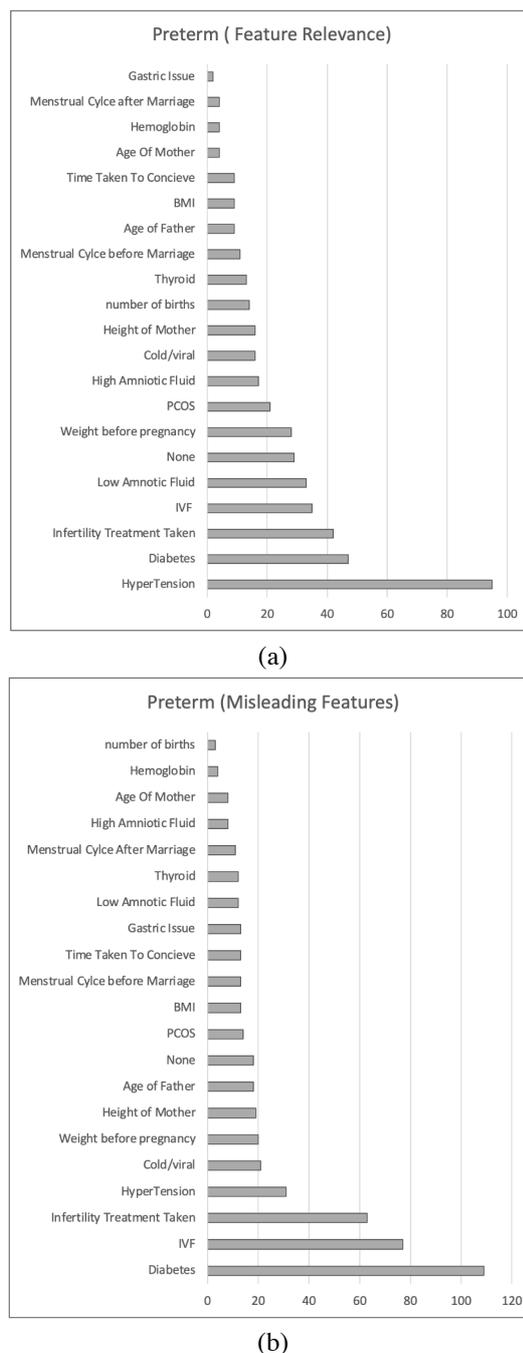


Fig. 1: Physical feature for Preterm birth prediction

Fig. 2(a),(b),(d),(e), shows the weights assigned to lifestyle features in terms of preterm childbirth predictions and NICU requirement prediction. Out of the 78 lifestyle features, few were never selected to be the

root nodes in best DF_Trees for the same reason figure shows weight of features selected as root nodes. Remaining features were considered to be having no

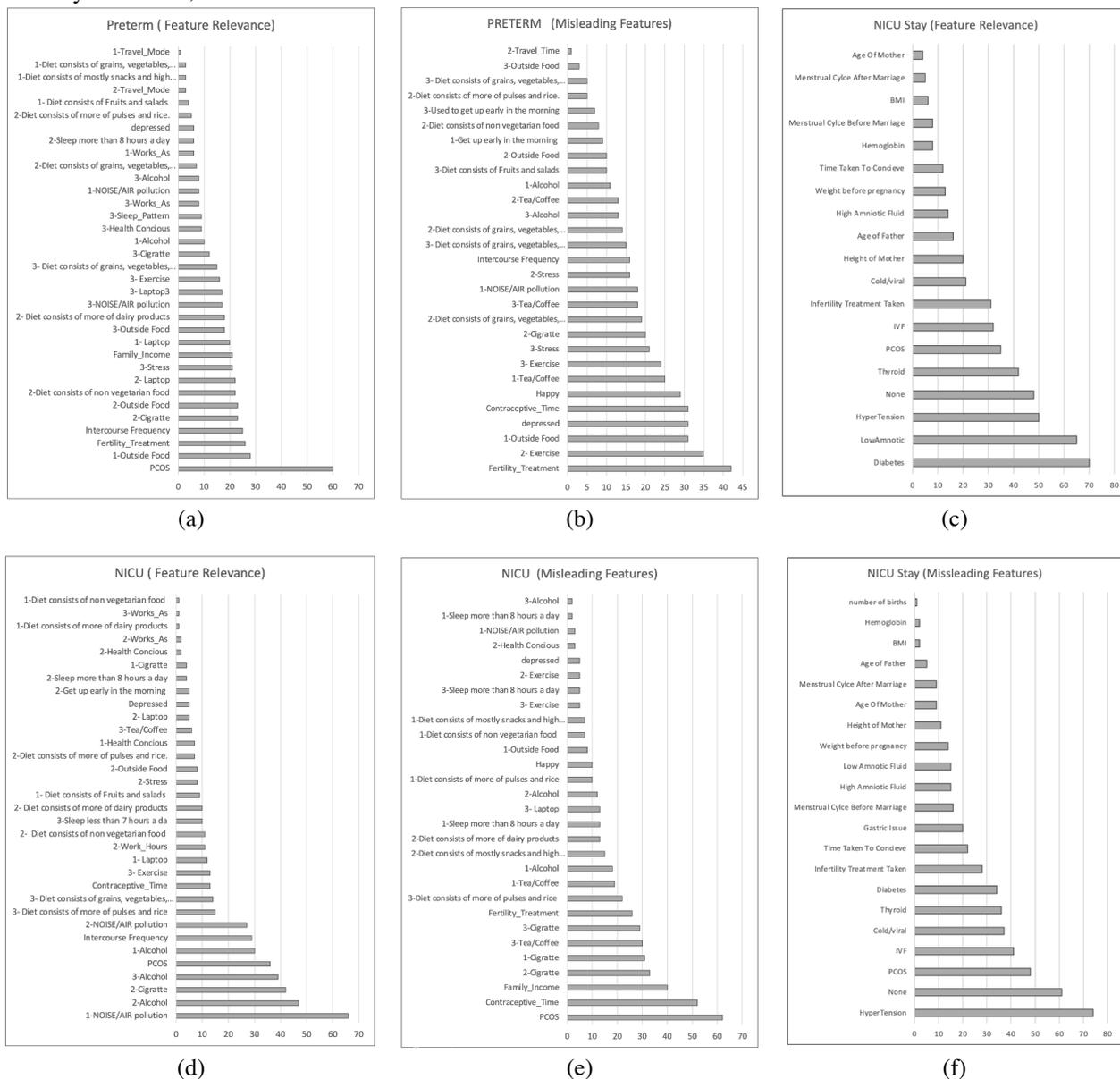


Fig. 2: Weight assigned to features (a),(b) lifestyle feature relevance and misclassification for preterm birth prediction, (d),(e) lifestyle feature relevance and misclassification for NICU prediction, (c), (f) physical feature relevance and misclassification for preterm birth prediction,

impact on prediction. Fig. 2(c),(f) shows the feature weight for relevance and misclassification of physical features for NICU requirement prediction. As the lifestyle features were spread across three phases of mother's life, in fig. (2), index 1 is used to depict teenage information, index 2 is used for after marriage information and 3 is used for information during pregnancy.

Looking at the weights assigned to physical features it has been observed that medical conditions like diabetes,

hypertension, infertility treatment, amniotic fluid are more relevant while predicting NICU requirement and preterm childbirth. Among the lifestyle features PCOS, consumption of alcohol, cigarette and outside food, noise and air pollution during teenage, usage of laptop after marriage, stress level during pregnancy are some of the relevant features observed after experiments. Features which are identified as misleading, create bias towards the majority class in imbalance dataset are IVF treatment, diabetes, usage of contraceptive, family income.

5 Conclusion

Woman reproductive health is getting effected with the change in her lifestyle across different phases of her life. A female needs to be educated about the impact these changes may have on her health. With this study we have used hybrid random forest model to analyse mother's physical and lifestyle features and the importance of these features in association with NICU requirement of child just after childbirth and possibility of preterm child birth complications. It has been observed that health conditions like diabetes and hypertension have an impact on pregnancy outcome, while her lifestyle habits like consumption of alcohol, unhealthy food also effects the pregnancy outcomes namely preterm birth and NICU requirement. Woman understanding the features that can impact her maternal health and child's health, can bring changes in her lifestyle thus avoiding such complications.

Future Work

We suggest researchers to explore more pregnancy complications like low birth weight, caesarian delivery, induction of pain etc. in relation with woman physical and lifestyle features, using different machine learning techniques towards better solutions.

References

- Series, R.D., Health Risk Factors in Rapidly Changing Economies.
- Ratner, R.E., Christophi, C.A., Metzger, B.E., Dabelea, D., Bennett, P.H., Pi-Sunyer, X., Fowler, S., Kahn, S.E. and Diabetes Prevention Program Research Group, 2008. Prevention of diabetes in women with a history of gestational diabetes: effects of metformin and lifestyle interventions. *The Journal of Clinical Endocrinology & Metabolism*, 93(12), pp.4774-4779.
- Goossens, J., Beeckman, D., Van Hecke, A., Delbaere, I. and Verhaeghe, S., 2018. Preconception lifestyle changes in women with planned pregnancies. *Midwifery*, 56, pp.112-120.
- Anand, A. and Shakti, D., 2015, September. Prediction of diabetes based on personal lifestyle indicators. In 2015 1st International Conference on Next Generation Computing Technologies (NGCT) (pp. 673-676). IEEE.
- Patil, M., Lobo, V.B., Puranik, P., Pawaskar, A., Pai, A. and Mishra, R., 2018, July. A proposed model for lifestyle disease prediction using support vector machine. In 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.
- Sharma, M. and Majumdar, P.K., 2009. Occupational lifestyle diseases: An emerging issue. *Indian journal of occupational and environmental medicine*, 13(3), p.109.
- Sharma, R., Biedenharn, K.R., Fedor, J.M. and Agarwal, A., 2013. Lifestyle factors and reproductive health: taking control of your fertility. *Reproductive biology and endocrinology*, 11(1), pp.1-15.
- Coutinho, E.D.C., Silva, C.B.D., Chaves, C.M.B., Nelas, P.A.B., Parreira, V.B.C., Amaral, M.O. and Duarte, J.C., 2014. Pregnancy and childbirth: What changes in the lifestyle of women who become mothers?. *Revista da Escola de Enfermagem da USP*, 48(SPE2), pp.17-24.
- Hemings, N., Greaves, L. and Poole, N., 2017. Preconception health care interventions: a scoping review. *Sexual & reproductive healthcare*, 14, pp.24-32
- Yi-Wei Chen and Chih-Jen Lin. Combining SVMs with Various Feature Selection Strategies. *Studies in Fuzziness and Soft Computing book series (STUDFUZZ, volume 207)*, 2006
- Desbordes Paul RuanSuModzelewski Romain V auclin Sebastien V era Pierre Gardin Isabelle. Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier. *Computerized Medical Imaging and Graphics*, Dec 2016
- Hanchuan Peng, Fuhui Long, and Chris Ding. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transaction on pattern analysis and machine intelligence* Vol. 27, No. 8, August 2005
- Chien-Pang Lee, Yungho Leu. A novel hybrid feature selection method for microarray data analysis. *Elsevier Journal Applied Soft Computing* 11 (2011) 208–213
- Hui-Huang Hsu, Cheng-Wei Hsieh, Ming-Da Lu. Hybrid feature selection by combining filters and wrappers. *Elsevier Journal Expert Systems with Applications* 38 (2011) 8144–8150
- P. Jaganathan, N. Rajkumar, and R. Kuppuchamy. A Comparative Study of Improved F-Score with Support Vector Machine and RBF Network for Breast Cancer Classification. *International Journal of Machine Learning and Computing*, Vol. 2, No. 6, December 2012
- Huijuan Lu, Junying Chen, Ke Yan, Qunjin, Yu Xue, Zhigang Gao, A Hybrid Feature Selection Algorithm for Gene Expression Data Classification, *Neurocomputing* (2017)
- Harun Uguz, A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm, *Knowledge-Based Systems*. Volume 24, Issue 7, October 2011, Pages 1024-1032
- Huilin Zheng, Hyun Woo Park, Dingkun Li, Kwang Ho Park, Keun Ho Ryu. A Hybrid Feature Selection Approach for Applying to Patients with Diabetes Mellitus: KNHANES 2013-2015. 2018 5th NAFOSTED Conference on Information and Computer Science (NICS)
- Himani Deshpande, Leena Ragha, April 22, 2021, "Mother's Significant Feature (MSF) Dataset", IEEE Dataport, doi: <https://dx.doi.org/10.21227/kq5k-b784>.
- Fang, L., Jiang, H. and Cui, S., 2017, July. An improved decision tree algorithm based on mutual information. In 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (pp. 1615-1620). IEEE.
- Deshpande H., Ragha L., (in press). A Hybrid Random Forest based Feature selection model using Mutual Information and F-score for Preterm birth classification., *International Journal of Medical Engineering and Informatics*.