Prediction of Over-the-Counter Products and Exercise as Home Treatment Modalities for Knee Joint Pain Using Deep Learning

Onengiyeofori Harry
Department of Computer Science
Texas Tech University
Lubbock, Texas
onharry@ttu.edu

Sruthi Mandalapu
Department of Computer Science
Texas Tech University
Lubbock, Texas
srmandal@ttu.edu

Abdurrhman Suliman
Department of Computer Science
Texas Tech University
Lubbock, Texas
absulima@ttu.edu

VenkataMohanaRao,Nandigam
Department of Computer Science
Texas Tech University
Lubbock, Texas
vnandiga@ttu.edu

Chandrika Kancheti
Department of Computer Science
Texas Tech University
Lubbock, Texas
ckanchet@ttu.edu

***Abstract*—Deep learning is utilized widely in the medical industry to detect and predict various outcomes. This paper focuses on utilizing a multilabel multiclass classification deep learning model to predict over the counter (otc) products, and exercises that a patient would use to treat knee pain. The project will be further extended to predict the user's post pain level after the selected treatment.**

***Index Terms*—multilabel multiclass classification, deep learning, synthetic data generation, knee pain**

I. INTRODUCTION

According to the National Library of Medicine [1] Knee joint pain affects approximately 25% of adults, and its prevalence has increased almost 65% over the past 20 years, accounting for nearly 4 million primary care visits annually. Although exercise and practice sport and reduce and delay knee pain, athletes who run or play sports that involve jumping or quick pivoting are also more likely to experience knee pain and problems [12]. Over time, it can be difficult to walk, run, use stairs, or engage in other activities requiring your legs. The good news is that several treatment options are available to reduce discomfort and help you get moving again. More than half of patients with knee joint pain utilize over the counter (OTC) pain medication. [2]. And Therapeutic exercise is often recommended as a means of treatment [3].

Machine learning algorithms have demonstrated potential in swiftly and efficiently predicting both the likelihood of patients acquiring specific diseases or conditions, and the most appropriate treatments. As a branch of computer science, machine learning utilizes algorithms to detect patterns within extensive datasets and aids in forecasting various outcomes [4]. In numerous disciplines, machine learning techniques have become key instruments for prediction and decision-making. The availability of clinical data has significantly enhanced the role of machine learning in medical decision-making [5]. The development of a machine learning model could provide essential support, enabling real-time, effective decisions regarding treatments or exercises.

In this paper we aim to develop a hybrid approach that combines multi label deep learning models to accurately predict the most effective over-the-counter (OTC) medications and exercises for knee joint pain, using data collected from patient surveys. This model will leverage advanced neural networks to analyze patient characteristics and treatment outcomes, enabling personalized treatment recommendations.

Specifically, we are going to address the following research questions in this work:

Can Deep learning algorithms be utilized to predict the OTC Products and Exercise as Home Treatment Modalities for Knee Joint Pain, and if yes, what is the predicted post pain level the subjects felt after the treatment?

The next sections cover a review of the techniques that will be utilized in this project, the methodology implemented and the results we derived from the experiments.

II. LITERATURE REVIEW

Typically to carry out a prediction would require data. At the time of the project that was unavailable and so the project was updated to two parts, first to generate synthetic tabular data upon which the data can be modelled, and secondly to build a model that can predict the treatments and post pain levels in one model.

We review methods for generating synthetic data, multilabel multiclass classification models for testing the efficiency of the synthetic data, and multi target deep learning model for predicting otc products, excersies and post pain levels.

A. Synthetic Tabular Data Generation

Data, and by extension synthetic data come in diverse types, from images to audio. For this project, the data is in tabular form. There are various methods for generating tabular data, but we review to main methods which can be divided into, the classical approaches and predictive approach [6].

The classical approach has two main methods

1) Baseline methods which utilize anonymization techniques and include "replacing data values, deleting sensitive attributes, and adding noise" [6] .
2) Statistical Models generate data by utilizing statistical and probabilistic models to simulate actual data and the relationships or correlations within the data [6].

The predictive approach also has two methods

1) Supervised Machine Learning Models which are trained on actual data and used to predict new records like the original records [6].
2) Deep Learning Approaches: In this case deep learning models are used to generate synthetic data. Two main methods include
    a) Autoencoders: An autoencoder comprises of an encoder, and decoder. The encoder compresses the data, creating and encoded representation of the real data, while the decoder decompresses the data so that it is a close variant of the original data [6].
    b) Generative Adversarial Networks comprise two neural networks, a generator and discriminator, that utilize an adversarial training process to generate synthetic data that is like real data [6].

Since actual data is unavailable in this project, we utilize the Statistical Method for generating the dataset utilized. The method is explained in the methodology.

B. Multilabel Multiclass Classification

Prediction models fall into two categories, Regression models which predict numerical outcomes, and Classification models which predict probability of an outcome belonging to a given class.

Typical prediction models have one target feature, for classification models these include binary classification where one output has two classes, or multiclass classification in which the output has more than two classes. Classes are exclusive categories.

Extending the problem set the classes of an output variable could be further categorized, that is labeled into non-exclusive categories, for example a bird can be white or blue, but it cannot be pigeon and a dove at the same time. This leads to the model domain of multilabel classification, and they can be broken into two groups single label classification and multilabel multiclass classification models.

Definition 1 (multi-label multi-class classification definition)**.** *Let X be a d-dimensional input space so that* $(x_1, x_2, \ldots, x_n) \in X$ *with, output space of q with L labels* $L = \{\lambda_1, \lambda_2, \ldots, \lambda_q\}$, *q>1 so that* $Y \subseteq L$ *is called a label-set. A multilabel dataset is defined by* $D = \{(x_i, Y) \mid 1 \leq i \leq m\}$. *Given a quality criterion Q that rewards high prediction and low complexity.*
A Multilabel Multiclass Classification aims to compute the function h:X$\rightarrow 2^1$ *such that h maximizes Q*

Multilabel Classification models can solve using the following methodologies

1) Problem Transformation Methods: These methods rely on transforming the dataset into a set of binary or multiclass problems, i.e., each dataset will have just one label. The main methods include
    a) Binary Relevance methods which transform the dataset in L binary classification problems, and [5]
    b) Label Powerset methods which creates a binary classifier for every label combination available in the dataset [7]
    c) Ensemble methods: Utilize a set of multi-class classifiers with methods like bagging and stacking to create a multi label ensemble classifier [5], [7].

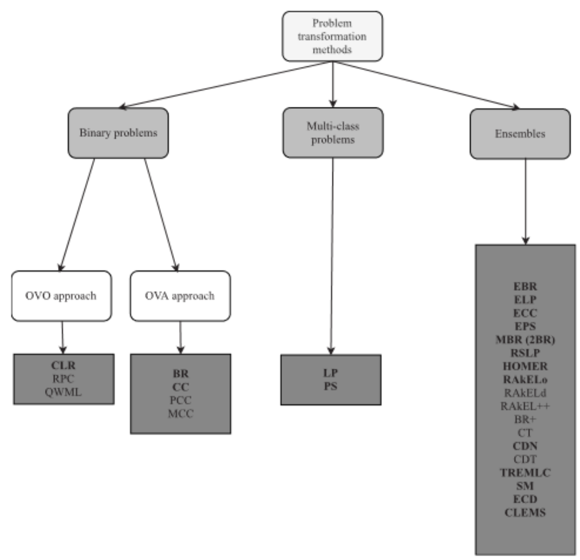

Fig. 1. Problem Transformation Methods [5]

2) Algorithm Adaptation Methods: They require adjustment to various machine learning and deep learning models so that instead of predicting one target variable, they can predict multiple. Figure II provides a description of such models and an explanation for an Artificial Neural Network follows
    a) Back Propagation Neural Networks (BPNN): This method utilizes a feed forward neural network with multiple output nodes in which each node represents a distinct label. It utilizes backpropagation to calculate the parameters of the network. Hyper parameters of the network are the learning rate, number of epochs and hidden units.Varying them has an impact on the performance of the

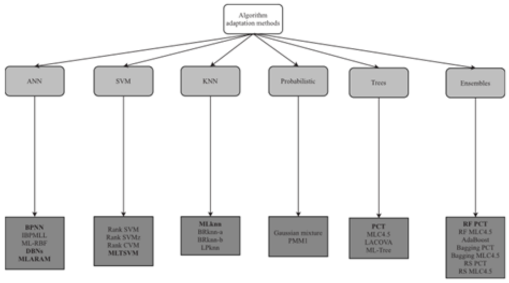

Fig. 2. Algorithm Adaptation Methods [5]

model. Our project utilizes a BPNN for testing the synthetic data generated which we illustrate in the methodology.

C. Multi-task Learning

Multi-task learning is an umbrella for a class of models that can predict multiple tasks see Figure 3, the idea being that a model can utilize training signals related to multiple related tasks to help improve generalization performance for those tasks [8]. A definition of multi-task learning is;

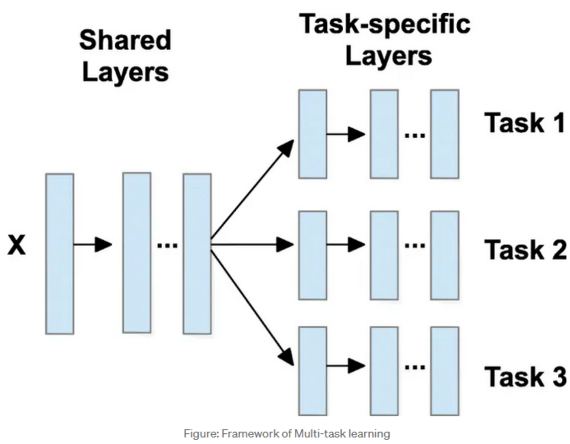

Fig. 3. Multi-task Learning [9]

Definition 2 (Multi-Task Learning)**.** *Let* $\{(\mathbf{x}_i^t, \mathbf{y}_i^t)\}_{i=1}^{N_t}$ *be the dataset for task* t*, where* $\mathbf{x}_i^t$ *is the input and* $\mathbf{y}_i^t$ *is the target output for the* i*-th sample of task* t*. In multi-task learning, we aim to learn a set of models* $\{f^t(\cdot)\}_{t=1}^T$ *for* T *tasks simultaneously, where each model* f^t *maps the input* $\mathbf{x}_i^t$ *to the output* $\mathbf{y}_i^t$*.*

The objective of multi-task learning is to minimize a combined loss function across all tasks, which can be expressed as:

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \lambda_t \sum_{i=1}^{N_t} \mathcal{L}_t(f^t(\mathbf{x}_i^t; \theta), \mathbf{y}_i^t)$$

where:

- $\mathcal{L}_t$ *is the loss function for task* t*.*
- λ_t *is the weight for task* t*, balancing its contribution to the total loss.*
- θ *represents the shared parameters across tasks, while task-specific parameters may also be included. [10]*

Two main factors that impact multi-task learning are:

1) how related tasks are, this needs to be encoded into the design of the model, and
2) how the task is defined. this is based upon the learning type, below are a list of the various types of multi-task multi-learning types;

1) multi-task supvervised learning: Task type can be homogenous or heterogenous, they could be all regression, or classification or both. Labels are available in training data for model to learn from [11].
2) multi-task unsupervised learning: These consist of clustering problems, in that the main task is to find grouping patterns within the data set. Train data are unlabeled [11].
3) multi-task semi-supervised learning: This is similar to the supervised learning, but this time some of the train data are unlabeled [11].
4) multi-task active learning: Tasks learn from unlabelled data by actively selecting unlabeled data instances to query their labels. [11].
5) multi-task reinforcement learning : Each task chooses actions that will maximize their reward [11].
6) multi-task online learning : In this case, tasks learn from streaming data [11].
7) multi-task multiview learning: Integrates multi-view learning into multi-task learning by utilizing multiple sets of features or views that describe the data, to learn to predict.

This paper focuses on multi-task supervised learning. As stated, supervised learning has labeled training data, and for multi-target learning (MTL), there are three primary models for learning;

1) Feature Based Multi-Target Learning (MTL): The feature based MTL learns common features for all tasks, where the feature representation can be subset or a transformation of original feature set /cite8.
2) Parameter Based Multi-Target Learning (MTL): This method uses coefficients of one task to learn the coefficients of other tasks. Methods include low rank approach, task clustering approach, task relation learning, and decomposition approach [12]
3) Instance Based Multi-Target Learning (MTL): This is a multi-task distribution matching method. It estimates the density ratio, between the probability of an instance (observation) and its label to determine wether it belongs to a task or a mixture of tasks /cite8

Since this project focuses more on utilizing deep learning models for the multi-task supervised learning, we expantiate more on the Feature Based Multi-Target Learning.

D. Feature Based Multi-Target Learning

The feature based approach relies on the fact that since tasks are related, there is features would have a representation that would be common to them. As such aim is to learn this feature representation that would provide model generalizations for the tasks. There are two major methods for acheiving this;

1) Feature Transformation Method: In this method the learned representation, is a transformation from the original features. An algorithm example relevant to our project is the multi-layer feed foward neural network. It can be expanded from having a single output, to having multiple outputs for various tasks as seen in Figure 4. This form of neural network has a regularization framework with an objective function defined in Definition 3

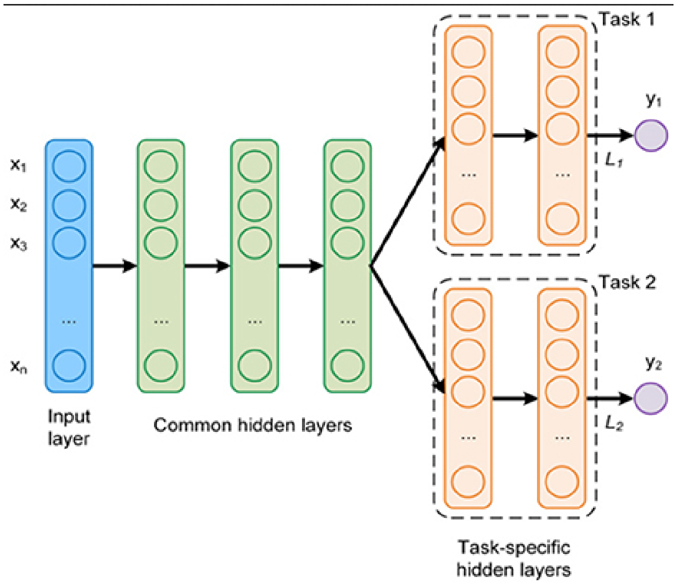

Fig. 4. Neural Network - Feature Transformation [13]

Definition 3.

$$\min_{\theta_s,\{\theta_t\}} \sum_{t=1}^{T} \alpha_t L_t\left(f_t\left(f_{shared}(X;\theta_s);\theta_t\right), Y_t\right) + \lambda\Omega(\theta_s,\{\theta_t\}) \tag{1}$$

where:

- T *is the total number of tasks.*
- α_t *is the weighting coefficient for task* t.
- L_t *is the loss function for task* t.
- $f_{shared}(X;\theta_s)$ *is the shared feature transformation, parameterized by* θ_s.
- $f_t(f_{shared}(X);\theta_t)$ *is the task-specific transformation, parameterized by* θ_t.
- Y_t *is the true output for task* t.
- λ *is the regularization parameter.*
- $\Omega(\theta_s,\{\theta_t\})$ *is the regularization term. [10]*

An alternative architecture is the cross-stitch network, which provides knowledge sharing and transfer between tasks. In this architecture, there are series of layer, with each layer including a cross-stitch unit. A cross-stitch unit takes input from shared layers and task specific layers, and their outputs are used as input for the next layers. Definition 4 provides a cross stich function in which the weight matrix and bias term are learned to combine shared features with task specific information [**?**]

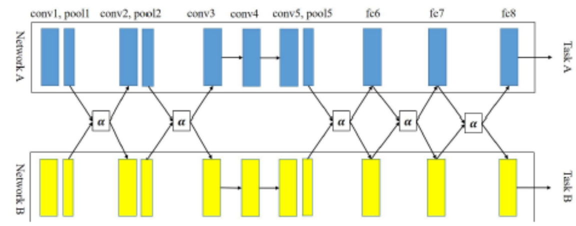

Fig. 5. Multi-task Learning [9]

Definition 4.

$$\mathbf{h}_i^t = \mathbf{W}_t\mathbf{h}_i^s + \mathbf{b}_t$$

where:

- $\mathbf{h}_i^s$ *is the vector of shared features for the* i*-th sample.*
- $\mathbf{W}_t$ *is the weight matrix specific to task* t*, which determines how much of the shared features are used for task* t.
- $\mathbf{b}_t$ *is the bias term specific to task* t.
- $\mathbf{h}_i^t$ *is the output of the cross-stitch unit for task* t.

2) Feature Selection Method: On the other hand, the feature selection method chooses a subset of the original features as the learned representation, thus eliminating features that do not have value for the tasks. Utilizing a subset of the original features also implies that the feature selection approach has more interpret ability compared to the feature transformation approach, since the features have been un-transformed. [8].

Definition 5 provides the definition of the objective function for the feature selection method. To select features, the regularizer minimizing $\|W\|_{2,1}$ forces it to become row-sparse, leading to all features with weights zero being shutoff.

Definition 5.

$$\min_{W,b} \; L(W,b) + \lambda\|W\|_{2,1} \tag{2}$$

where:

- $L(W,b)$ *is the loss function associated with the model predictions.*
- W *is the weight matrix with dimensions* $d \times T$*, where* d *is the number of features and* T *is the number of tasks.*
- b *is the bias term for each task.*
- λ *is the regularization parameter that controls the trade-off between the loss and the regularization term.*
- $\|W\|_{2,1}$ *is the mixed-norm regularization term, defined as the sum of the* ℓ_2*-norms of the rows of* W:

$$\|W\|_{2,1} = \sum_{i=1}^{d} \|W_{i,:}\|_2$$

where $\|W_{i,:}\|_2$ *is the* ℓ_2*-norm of the* i*-th row of* W.

III. METHODOLOGY

Our problem requires building a multi-task deep learning model on synthetic data. The problem definition thus required that we split the methodology into two phases where Phase 1 would be to understand the data model, how to represent the dataset and build a baseline dataset and neural network model that we can test on. Phase 2 would then involve the actual construction of the dataset which would then be tested on a multilabel multiclassification models to determine the efficiency of the dataset and finally a multi-target deep learning model to predict selected otc products, accompanying exercises and post pain level from their application.

A. Phase I

1) Baseline Synthetic Data: While actual data was not made available, a survey template was provided from which we got the data model. The structure of the template has 91 features which are a mix of numerical and categorical data, and 6 output variables which can be broken into two groups – OTCSelect and ExcersiceSelect. Each OTCSelect question has 32 classes, while each ExcersiseSelect question has 53 classes.

From these we constructed the baseline data model as having 6 labels, 85 classes and 264 features. The output features represent all the classes, and each class can have a probability of 0 to 1, for each label assigned to each class. In counting the number of features, it was necessary to include a count of the classes of the categorical variables since they would be dummy coded. This gives us the result that data could be high dimensional for both X and Y, if the number of features is much greater than the number of observations. We decided to experiment with two types of synthetic datasets, high dimensional dataset, or wide dataset, and a long dataset in which the number of observations is much more than the number of features.

Utilizing the data model above, we generated the dataset using the "make_multilabel_classification" function from sklearn.datasets. This gives us the baseline dataset to test a baseline model.

2) Baseline Multilabel Multiclass Classifier: For the model we utilize a dense 4-layer BPNN. A picture of the layers is provided Figure 4. We utilize RELU as activation for the input and hidden layers, Sigmoid activation function as activation for the output label as explained above the sigmoid function ensures that each label has a probability between 0 to 1 that is assigned to each class. Binary cross entropy is utilized for the loss function.

B. Phase II

1) Final Synthetic Data: Synthetic Data: In this phase we begin the actual construction of the data set using the provided data model. Since we do not have any actual data, we can derive information from past research on knee pain, and construct data based on the conditional probabilities either provided or that we can derive from research. Fig provides a flow chart for data generation.

First, we divided the dataset into two parts the demographic information and the knee pain survey data. The demographic data was sourced from freely available public data [14], [15]. The knee pain data, on the other hand, is not freely available. For that part we rely on provided information from past research to determine what fraction of our population should be positive for a specific condition. For example, to determine the cause of knee pain "OTCCause" we find from past research that some causes of knee pain are aging, and obesity [16], we also determine that a BMI above 25 is overweight [17], with that we can set all samples with BMI greater than 25 as cause obese, and all samples with age greater than 50 as aging.Figure 6 provides an example where based on an observation's JobTitle, Education, and Employment, the observation is categorized as either having the job categoy "White_Collar" job or "Blue_Collar" job.

Fig. 6. Sample Data Generation Process []

Utilizing data from research in building out the features from the samples also helps us in having in-built relationships which can be modeled.

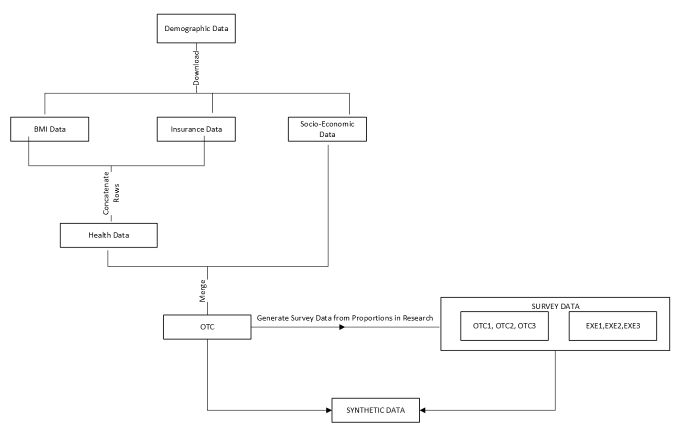

Fig. 7. Synthetic Data Generation Map []

2) Final Multilabel Multiclass Classifier: To test the efficiency of the dataset, we implement a dense 4 layer BPNN model, with 1 input layer, 4 hidden layer and 1 output layer. Figure 7 is a rendering of the model. Again the activation function for classifying the labels is the sigmoid function, with binary crossentropy utilized as the loss. The train/test set is split 80/20.

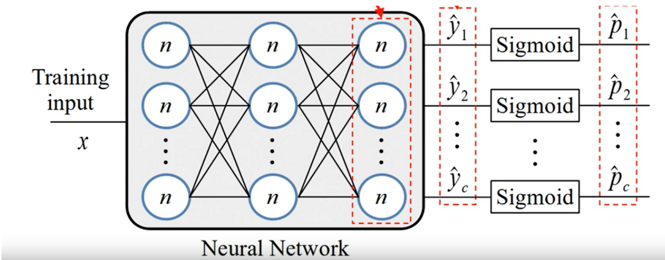

Fig. 8. Sample Multi-Label Multi-Class Model []

C. Multi-target Deep Learning Model

For the multi-target we utilize a 2-node neural network, based on the idea of feature transformation, discussed in literature review. We created a model akin to Figure 3, with a common input layer, hidden layers and seperate outputs for the various tasks. We made an adjustment to the PostPain variable and converted it to categorical embedings following mapping;

1) 0 - 2: Low Pain
2) 3 - 5: Medium Pain
3) Above 5: High Pain

Following the general model outlined, we test the model by changing, number of hidden nodes for the common layer, and the task specific layer to see how it impacts accuracy.Figure 9 provides the tensorflow summary of the model built

Layer (type)	Output Shape	Param #	Connected to
input_layer_9 (InputLayer)	(None, 235)	0	-
dense_86 (Dense)	(None, 235)	55,460	input_layer_9[0][0]
dense_87 (Dense)	(None, 200)	47,200	dense_86[0][0]
dense_88 (Dense)	(None, 200)	40,200	dense_87[0][0]
dense_89 (Dense)	(None, 200)	40,200	dense_88[0][0]
dense_90 (Dense)	(None, 196)	39,396	dense_89[0][0]
dense_91 (Dense)	(None, 17)	3,417	dense_89[0][0]

Fig. 9. MTL Model Summary []

IV. Results and Discussion

The results are broken down by the results from the phases

A. Phase I

In this section we inspect the impact of the generic synthetic data sets on the built model.The table below shows the results for both the datasets. Accuracy is used to measure the model's performance. From the results we infer from it that a neural network would work well in making predictions on the data model even if it is a high dimensional dataset. More critically we would need to ensure that the dataset built has relationships/correlations that can be modelled.

Size of Dataset	Accuracy
Long Dataset	92.8%
Wide Dataset	93%

TABLE I
Accuracy over Different sizes of dataset

B. Phase II

For this section we look at the proportions generated from the synthetic dataset, the result from the multilabel classification model, and finally the multi-target model.

1) Final Synthetic Data: The utilization of actual health and socio-economic data for generating demographic data provides a so-called "ground-truth" data upon which the rest of the survey data can be generated from. We start with an exploratory analysis of the data to see how the proportions compare to information gathered from research

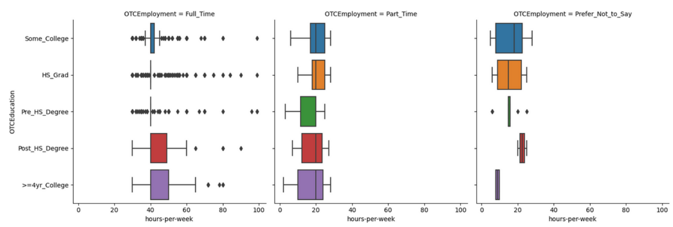

Fig. 10. Work Hours By Employment and Education

Figure 10 is of the number of hours people in the data set work, grouped by their education level and their employment status. First of all it can be seen that people with at least a high school degree tend to have full time jobs and are less likely to be ambiguous about their employment status. Also from research we can also infer that workers with full time degrees are less likely to work for organizations with mandates for the number of hours they have to stand [18]. This in turn goes on to affect how long a person in the dataset likely to stand for (OTCStand) based on their education and employment status as seen in Figure 11. From this we can see that workers with less than a high school degree are more likely to stand for longer hours a work.

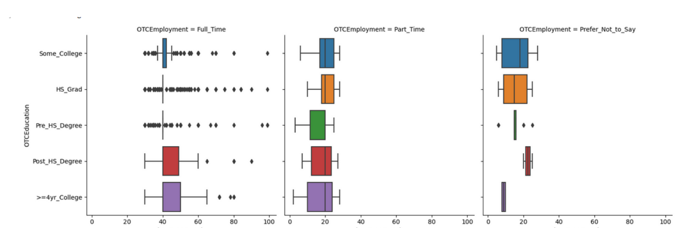

Fig. 11. Hours Standing By Employment and Education

Another example which is more common with knee pain is Obesity, in Figure 12, bmi data in the demographic data can be used to infer if a person has ever done excersise (OTCEverDoneExcersise), with all users with bmi greater than or equal to 25 set to no and less than 25 set to yes.

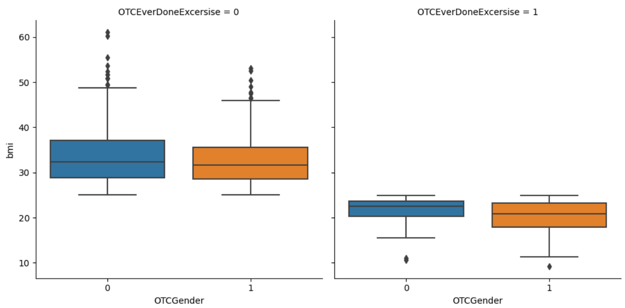

Fig. 12. Ever Excersised: BMI by Gender

2) Multilabel MultiClass Model: The efficiency of the finaly synthetic data can only be estimated when applied to a model. Note that this does not equate to the utility of the dataset as that can only be tested when compared to an actual real world dataset. Since that is unavailable we rely on testing the generated data set on a multi-label multi-class model as a test of the efficiency of the dataset for two reasons;

1) A dataset with purely random values cannot be modelled as there is no underlying distribution from which the data values where generated. In other words the probability for each value 0.5 and so between features if a linear correlation would be 0 because there no underlying covariance. This is one challenge our data generation method seeks to overcome, and any model should be able to model some of the variation in the dataset.
2) We utilize a multi-label multi-class model because as we have tested from the Phase 1 with a correlated data model, it will be capable of modelling the data if indeed the data has the underlying covariance as discussed in item 1.
   Figure 13 is a comparison of the accracies between this final model and the baseline models.This final model gives a test accuracy of 97%. Figure 14 gives the training and validation loss from which we see that the model is able to generalize quite well on the dataset. This means that the dataset can be utilized for building a model.

Type	Sizeof Dataset	Test Accuracy
Baseline	Long Dataset	92.80%
Baseline	Wide Dataset	93%
Final	Long Dataset	**97%**

Fig. 13. Comparison of Model Test Results

3) Multi-Target Model: Since the multi-target learning model learns for two tasks it returns two results as seen Figure 15. From the results, the model is able to learn the task for selecting OTC product and accompanying exercise. Since the post-pain is converted to categories before modelling, the result is also an accuracy of 74%. This value also tells us that the Post Pain levels in the synthetic dataset are not random and so more feature engineering and model tweaking should produce even better results.

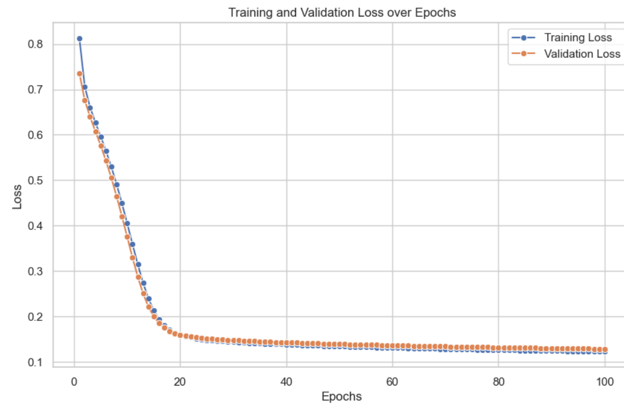

Fig. 14. Train and Validation Loss

Type	Test Accuracy
OTC &Exe	96.90%
PostPain	74%

Fig. 15. Multi-Target Learning Test Accuracy

V. Future Directions

From further research we came to understand that we could utilize a loss function separately for the PostPain variable and keep it as a regression. Doing this would improve model prediction especially if the categorical variable is kept as a feature, which can provide meaningful training signal.

The Demographic data and OTC1 survey data where hand coded, with the remaining survey data replicated from OTC1 survey and Execersie outcome variables changed to the required excersises. Utilizing a Generative Adversarial Model, the remaining OTC and Excersise survey data can be generated/predicted by using the Demographic and OTC1 survey data as the ”ground-truth’ data.

Some form of feature selection must be implemented so that variables with little to no information can be filtered out from the final feature set of the data. In so doing the models will generalize better because of the noise removal.

VI. Conclusion

This project investigaged the use of statistical models for generating synthetic datasets. The demographic part of the dataset was sourced from publicly available databases and used to construct the synthetic dataset, with the survey part of the dataset generated using proportions from research. Modelling the dataset using Multilabel multiclass models indicate that the dataset is efficient. It was further used on a multi-target learning model which produced an accuracy of 97% for the OTC product task, and 73% for the post pain task. We conclude that the synthetic dataset is efficient for carrying out modelling tasks in lieu of actual data.

ACKNOWLEDGMENT

We are thankful to Dr. Sheng for providing us the opportunity to work on his research project.

REFERENCES

[1] Machine learning technique. https://www.sciencedirect.com/topics/computer-science/machine-learning-technique.

[2] S. J. M. Merchant, A. C. Li, A. P. Nguyen, M. N. Wirtzfeld, A. E. McLeod, J. S. Park, D. P. Dixon, J. N. Bathe, and A. B. Ball. Clinicopathologic factors and adjuvant therapy in colonic perforation. *Canadian Journal of Gastroenterology and Hepatology*, 2016:Article ID 1360345, 2016.

[3] C.-C. Wu, W.-C. Yeh, W.-D. Hsu, Md. M. Islam, P. A. Nguyen, T. N. Poly, Y.-C. Wang, H.-C. Yang, and Y.-C. (J.) Li. Prediction of fatty liver disease using machine learning algorithms. *Computer Methods and Programs in Biomedicine*, 170:23–29, 2019.

[4] R. Christensen, E. M. Bartels, A. Astrup, and H. Bliddal. Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: A systematic review and meta-analysis. *Annals of the Rheumatic Diseases*, 66(4):433–439, 2007.

[5] J. Bogatinovski, L. Todorovski, S. Džeroski, and D. Kocev. Comprehensive comparative study of multi-label classification methods. *Expert Systems with Applications*, 2022.

[6] M. Hernandez, G. Epelde, A. Alberdi, R. Cilla, and D. Rankin. Synthetic data generation for tabular health records: A systematic review. *Neurocomputing*, 2022.

[7] Multi-label classification. urlhttps://en.wikipedia.org/wiki/Multi-label$_{c}lassification.Accessed:$ $2024-07-31$.

[8] Yu Zhang and Qiang Yang. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(1):558–576, 2021.

[9] Rohan Sharma. Multi-task learning: All you need to know (part 1). *Python in Plain English*, Dec 2022.

[10] OpenAI. Chatgpt. Online, Jul. 31 2024. Available: https://chat.openai.com/.

[11] Yu Zhang and Qiang Yang. An overview of multi-task learning. *National Science Review*, 5(1):30–43, 2018. Advance access publication 1 September 2017.

[12] A. N. Tarekegn, M. Ullah, and F. A. Cheikh. Deep learning for multi-label learning: A comprehensive survey. *IEEE Transactions on Neural Networks and Learning System*. To be published.

[13] Author(s) Full Name. Full article title. *Frontiers in Artificial Intelligence*, Volume Number(Issue Number):Article Number, 2024.

[14] Describing relationships between two variables. https://online.stat.psu.edu/stat200/book/export/html/242.

[15] Healthcare insurance dataset. https://www.kaggle.com/datasets/willianoliveiragibin/healthcare-insurance.

[16] Knee pain - symptoms and causes. https://www.mayoclinic.org/diseases-conditions/knee-pain/symptoms-causes/syc-20350849.

[17] The impact of obesity on bone and joint health. https://www.aaos.org/contentassets/1cd7f41417ec4dd4b5c4c48532183b96/1184-the-impact-of-obesity-on-bone-and-joint-health1.pdf.

[18] U.S. Bureau of Labor Statistics. Occupational requirements survey. *U.S. Department of Labor*, Jul 2024. https://www.bls.gov/news.release/ors.nr0.htm.

[19] The impact of obesity on bone and joint health. https://www.aaos.org/contentassets/1cd7f41417ec4dd4b5c4c48532183b96/1184-the-impact-of-obesity-on-bone-and-joint-health1.pdf.

[20] Management of hypertriglyceridemia: Common questions and answers. https://www.aafp.org/pubs/afp/issues/2018/1101/p576.html.