Read the below paper and write a report limited to 5 pages, including citations, figures etc. Reports beyond 5 pages will be rejected automatically.

The report should contain whether you think doppelganger effects are unique to biomedical data, and how you think it can be avoided in the practice and development of machine

learning models for health and medical science.

Answer:

INTRODUCTION:
Machine learning a sub concept of artificial intelligence . Machine learning involves various algorithms and statistical model which makes a larger dataset to be delivered in simpler and explainable form to others.
Now a days machine learning in biomedical field has become very important .Here the large data are analysed using different algorithms like classification, regression and clustering .
Doppel ganger effect is nothing but different concepts and objects appears identical . We are seeing this effect now in both biomedical science side and in data science as well.
First let's see about doppelganger effect in biomedical science , here it occurs with biological samples like, blood, urine and cell , sometimes with patients also with features like genetic markers, imaging data like x rays, MRI, ultrasounds and clinical symptoms as well, i.e. the biological samples and the patients heath screening features looks identical .
Now let's see doppelganger effect in data science, here this occurs when the different models and algorithms of a machine learning concept shows similar results .
Patients through clinical analysis is done by different respective equipment's by specialised doctors , but all these records are saved as a data . This data is forwarded to machine learning where first we are exploring the structure of the data , then our very next step is the Exploratory data analysis or cleaning of data , then its data prediction splitting into test and train data , then there are two branches whether to use supervised machine learning or unsupervised machine learning . Supervised machine learning uses regression and classification . Regression uses algorithms like linear regression, polynomial regression and K nearest neighbours, Classification uses algorithms like Naïve bayes theory and many more . Unsupervised learning uses unlabelled dataset using two main types i.e. clustering and dimensionality reduction. Clustering uses algorithms like k means clustering, hierarchical clustering and DB scan .Dimensionality reduction uses algorithms like principal component analysis and t scan .
Doppelganger effect is addressed by cross validation, regularization and ensemble learning which reduces overfitting of the model and gives generalized performance .
In this study we are using Doppelganger effect in patients clinical record analysis. This study explains the Doppelgangers uses and their draw backs in the patients clinical results . How is it unique in this biomedical data , what haywires and what practices is done to reduce that effect in biomedical data and in Machine learning .
When you tell about uniqueness , it should be distinct from others , i.e. it should not be used in a routine and should be used selectively or rarely . Doppelganger effect is something new , but when we are seeing the uniqueness towards biomedical data it's a big no . Doppel gangers effect is seen in other biomedical data types like proteomics, metabolomics , imaging, genomics and microbiome .
In genomics data we get Gene sequencing samples ,this effect is shown in two genetic variants which shows similar nucleotide sequencing , In imaging data we can take x ray scans

or MRI these shows X ray reports of two different patients ,where the structure shown for both is similar , In proteomics data ,the patients take a report of their protein identification mainly the peptide sequence ,here their physical appearance looks similar .
Doppelgangers effect has been in action widely among different data types so it is not unique for biomedical data .
Do you think doppelganger effects is helpful or not helpful in machine learning in health and data science :
In my point of view Doppel ganger effect shows similar data in respect to other data which leads to inconsistency of the machine learning process, i.e. The final result after the train and test data prediction would be wrong , error in diagnosis . So it wont be helpful for analyses of data to get an accurate answer. But it is helpful if we carefully evaluate and develop some practices to overcome this errors .
Now lets see how to avoid or mitigate doppel gangers effect while practicing and developing machine learning in biomedical data ,They are
Data diversity , that is the training data will have a wide range of samples where we can avoid different individuals having similar profiles ,by collection of data from different demographics, geographies and medical condition. (Wu et al 2019)
Unique selection , selection of required and important features which ultimately differentiates between persons , that is more of data cleaning which gives a precise data ,this is done with the help of statistical methods .
Updating the machine learning models frequently to feed in new data and avoiding recycling of old data , this gives clearer minds to the machine to identify and reduce doppel gangers effect .
The most scientific and elaborate method is usage of algorithm techniques . when we get to know we are using the similar technique for similar data we can get to know of the doppelganger effect and avoid it . The techniques used are clustering and outlier detection.
Lastly Human evaluation, machines also will have errors, we humans should be careful while development and deployment of models made using machine learning .
If all these techniques are watched and analysed throughly doppel ganger effect error will be reduced.

Conclusion:
Machine learning is vast and one of the latest learning to summarise and understand large data's , Doppelgangers effect shows errors in analysing resultant data but the practices developed can overcome this effect .

REFRENCES:
Machine Learning in Biomedicine : A review of techniques and applications " by Ghassemi et al (2019).
Rajkomar,A., Dean, J., & Kohane,I.(2019).Machine learning in medicine . New England Journal of Medicine, 380(14), 1347 – 1358
Saria et al (2019). Review highlights need for rigorous validation procedures to ensure accuracy and effectiveness of machine learning model .