Predicting Idiopathic Pulmonary Fibrosis Progression

Arnav Kumar
Grade 11 Webber Academy Applied Science Student*

Correspondence to **Dr. Christian Jacob**
University of Calgary - Department of Computer Science,
Department of Biochemistry and Molecular Biology

Introduction

Idiopathic Pulmonary Fibrosis. Idiopathic Pulmonary Fibrosis (IPF) or Cryptogenic Fibrosing Alveolitis (CFA) is a disease affecting the lung base and leads to lung function decline with little to no therapies available other than lung transplant (*1, 2*). Although it was previously believed that the disease affects only 5 out of every 100,000 individuals, the disease is now known to be much more prevalent (*1, 3, 4*). The disease is age-related but does not have any known cause and mainly affects older patients with the median age at diagnosis being 66 (*4, 5*). Recently, there have been claims that it is a result of abnormally activated alveolar epithelial cells (*5*). Patients experience a shortness of breath, and some features of the disease include diffuse pulmonary infiltrates recognizable by radiography and varying degrees of inflammation or fibrosis (*2*). Affected lung areas alternate with unaffected areas in the lung (*2*). Affected areas are characterized by the differences in cell age and due to a honeycomb fibrosis pattern (*2*).

The outcome of Pulmonary Fibrosis can range from rapid health declination to a healthy stability, but doctors are unable to easily diagnose the severity of the disease. There exist meth-

*Under Supervision of Dr. Garcia-Diaz

ods to diagnose severity, but these can be complicated and are not standardized (*6*). An example of such a method is a cough scale questionnaire or a shortness of breath questionnaire (*6–8*). Another method of diagnosing severity is through a functionality test known as the 6 month 6 minute Walk Distance or 6MWD test, but as the name suggests, this test is not instantaneous, and still requires the effort of trained professionals (*6, 9*). On the other hand, Machine learning has been used with data from different points in time to provide a prognosis by using a software tool called CALIPER that uses radiological changes to predict IPF severity (*10*). Another case of using machine learning used computed tomography (CT) scans of the lung region and obtained an accuracy of around 76.4% or 70.7%, only outperformed 66% of doctors and only classified the severity rather than providing numerical estimates (*11*). An accurate prognosis of the disease will put patients at more ease, and may pave the path for any treatments that will come in the future. For this reason, it is essential that a consistent and easy method for diagnosing the severity of the disease is found.

Deep Learning Methods. Machine learning is a good fit for the task at hand because doctors can let the program run given the data, and it has been used in the past to diagnose other diseases and make predictions (*12*). Although machine learning has been used before for this task (*6, 9, 10*), the accuracy of the models can be improved on. Furthermore, a machine learning model could make it easier to get a prognosis.

For a disease such as IPF which is a fibrosing disease within the lungs, imaging the lungs through CT scans yields in enough insight to accurately evaluate the patients prognosis (*13*).

Furthermore, for injuries like neck fractures, machine learning has proven to be an improvement to the prediction performance using a method of bayesian classification (*14*). For diseases like cancer, machine learning has also been used to give a prognosis and modern machine learning methods have been shown to outperform more classical methods including decision trees (*15*). On another note, machine learning has already been used with images of leafs to determine plant diseases and their severity, showing the ability to handle and diagnose disease severity based on a CT scan input using machine learning (*16*).

Question This study aims to create a model that uses one baseline CT scan, as well as the forced vital capacity (FVC) of the lungs over the time period of one to two years. The model then predicts the FVC of the lungs for the next 3 checkups, and hence predicting the rate at which the lung condition degrades. The main questions of interest are: what is the greatest accuracy a machine learning model can attain in predicting the FVC of a IPF patient on their next 3 checkups, and what method produces this accuracy?

References and Notes

1. R. J. Mason, M. I. Schwarz, G. W. Hunninghake, R. A. Musson, *American Journal of Respiratory and Critical Care Medicine* **160**, 1771 (1999). doi:10.1164/ajrccm.160.5.9903009.

2. T. J. Gross, G. W. Hunninghake, *New England Journal of Medicine* **345**, 517 (2001). doi:10.1056/NEJMra003200.

3. D. B. Coultas, R. E. Zumwalt, W. C. Black, R. E. Sobonya, *American journal of respiratory and critical care medicine* **150**, 967 (1994). doi:10.1164/ajrccm.150.4.7921471.

4. G. Raghu, *et al.*, *American journal of respiratory and critical care medicine* **198**, e44 (2018). doi:10.1164/rccm.201807-1255ST.

5. T. E. King Jr, A. Pardo, M. Selman, *The Lancet* **378**, 1949 (2011). doi:10.1016/S0140-6736(11)60052-4.

6. H. Robbie, C. Daccord, F. Chua, A. Devaraj, *European Respiratory Review* **26** (2017). doi:10.1183/16000617.0051-2017.

7. T. E. King Jr, *et al.*, *New England Journal of Medicine* **370**, 2083 (2014). doi:10.1056/NEJMoa1402582.

8. M. J. van Manen, *et al.*, *European Respiratory Review* **25**, 278 (2016). doi:10.1183/16000617.0090-2015.

9. R. M. du Bois, *et al.*, *European Respiratory Journal* **43**, 1421 (2014). doi:10.1183/09031936.00131813.

10. F. Maldonado, *et al.*, *European Respiratory Journal* **43**, 204 (2014). doi:10.1183/09031936.00071812.

11. S. L. Walsh, L. Calandriello, M. Silva, N. Sverzellati, *The Lancet Respiratory Medicine* **6**, 837 (2018). doi:10.1016/S2213-2600(18)30286-8.

12. Y. Wang, Y. Fan, P. Bhatt, C. Davatzikos, *Neuroimage* **50**, 1519 (2010). doi:10.1016/j.neuroimage.2009.12.092.

13. S. L. Walsh, *et al.*, *European Respiratory Review* **27** (2018). doi:10.1183/16000617.0073-2018.

14. M. Kukar, I. Kononenko, T. Silvester, *Artificial intelligence in medicine* **8**, 431 (1996). doi:10.1016/S0933-3657(96)00351-X.

15. J. A. Cruz, D. S. Wishart, *Cancer informatics* **2**, 117693510600200030 (2006). doi:10.1177/117693510600200030.

16. E. Mwebaze, G. Owomugisha, *2016 15th IEEE international conference on machine learning and applications (ICMLA)* (IEEE, 2016), pp. 158–163. doi:10.1109/ICMLA.2016.0034.