# Prediction of the $g$ factors of TMD heterobilayers using Machine Learning Methods.

**Karol M. Gałązka[1], Maciej E. Marchwiany[2], T. Woźniak[1] and M. Birowska[1]**

[1] *University of Warsaw, Faculty of Physics, Pasteura 5, 00-092 Warsaw, Poland*
[2] *Interdisciplinary Centre for Mathematical and Computational Modelling (ICM), University of Warsaw, Pawińskiego 5a, 02-106 Warsaw, Poland*

Continued progress of technology and industry necessitates the development of tailor-made materials for specific purposes. In recent years the vast class of two-dimensional (2D) van der Waals (vdW) layered materials are extensively studied. Among them, Transition Metal Dicalchogenide (TMD) semiconductors have been subjects of particular interest in electronics and optoelectronics. [1]

In principle desired properties could be obtained by vertically stacking various types of monolayer (ML) crystals with arbitrary material sequence and twist angle between layers, owing to their synergetic effects. [2,3] Although, the interest in this pursuit is continuously increasing, the lack of high-throughput techniques for fabricating high-quality vdW heterostructures (HTs) hinders its research. The speed of the process could be expedited with the use of theoretical calculations to predict promising structures with desired properties. The $g$ factor that quantifies the response of excitons to external magnetic field is one of such a property, that has been widely studied in the framework of the density functional theory for MLs of TMDs. [3] However, $g$ factor research for HTs are limited due to requirement of large number of conduction bands to achieve convergence. [3] Additionally, the stacking layers and twist angles dependence further complicate the research.

This high computational demand and labor intensive process of prediction limits the pace of progress in the field. Many TMD monolayers phenomena have been extensively examined in relation to the $g$ factor. Namely, experimental studies of exciton complexes, excited Rydberg states etc., and theoretical computations based on electronic structure features. The idea is to investigate and utilise their predictive power using Machine Learning Methods (MLMs), applying this knowledge to TMD vdW HTs.

The first step in this process was data preparation, where all available data was collected and considered. It was then cleaned, all damaged or uncompleted records were removed or completed and initial descriptors (features) of records were established. Next a variety of Decision Tree Regressor based algorithms were fitted to the data, establishing the predictive power of each feature and yielding satisfactory fits to the data. The feature set was then progressively thinned, removing features of least importance and a variety of models were fitted to the heterobilayers data. The best model was selected based on calculated fit metrics, establishing decent predictive power for the $g$ factor.

[1] J. Qi, Z. Wu, W. Wang, K. Bao, L. Wang, J. Wu, C. Ke, Y. Xu and Q. He, *Int. J. Extrem. Manuf.* **5**, 022007 (2023).

[2] K. S. Novoselov , A. Mishchenko, A. Carvalho, and A. H. Castro Neto, *Science* **353**, 6298 (2016).

[3] T. Woźniak, P. E. Faria Junior, G. Seifert, A. Chaves, and J. Kunstmann, *Phys. Rev. B* **101**, 235408 (2020).