# PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints

* PaDEL-Descriptor is a software for calculating molecular descriptors and fingerprints. It calculates 797 descriptors and 10 types of fingerprints.
* Methods: - PaDEL-Descriptor was developed using the java language and consists of a library component and an interface component. The software uses a Master/Worker pattern to take advantage of the multiple CPU cores that are present in most modern computers to speed up calculations of molecular descriptors.
* Introduction: - a molecular descriptor “is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment”.
* A good descriptor calculation software should posses most of the following features:
  + Free or cheap to purchase so that is easily available to researchers.
  + Open source so that researchers could add in their own descriptor calculation algorithm
  + Has a graphical user interface (GUI) for easy usage and a command line version to allow the software to run in computer clusters through a software to run in computer clusters through a software job scheduler.
  + Able to work on multiple platforms such as Windows, Linux, and MacOS.
  + Able to calculate many types of descriptors.
* Methods: - the PaDEL-Descriptor has two component- a library component and an interface component. The library component is self-contained and it can be easily integrated into other QSAR software to provide the descriptor calculation feature.
* It provides wrapper classes around the 43 molecular descriptor algorithms and seven fingerprint algorithms that were implemented in CDK.
* The library component also included 4 molecular descriptors and three fingerprint algorithms. These include atom type electrotopological state descriptors, McGowan volume, molecular linear free energy relation descriptors, ring counts, count of chemical substructures identified by Laggner, and binary fingerprints and count of chemical substructures identified by Klekota and Roth.
* Calculation of molecular descriptors is performed in parallel by using a Master/Worker pattern, it has one master thread and one or more worker thread.
* The master thread starts the calculation process by reading molecular files and creates a job description for each molecule.
* A job description consists the name and structure of the molecule, preprocessing tasks that are to done before calculation of descriptors, and the types of descriptors and fingerprints to calculate.
* The jobs are added to a shared job queue and each worker thread will retrieve a job from the shared job queue and calculates the descriptors and fingerprints for the molecule specified in the job descriptors.
* The Master thread also provides functions to set the type of preprocessing tasks to be done, the type of descriptors and fingerprints to calculate, and the number of worker threads.
* The interface component provides both GUI and command line interfaces for operating and the software. It provides an easy-to-use interface for users to set the various options and select the individual types of descriptors and fingerprints to calculate.
* The command line interface was implemented using the Apache Commons CLI library.
* Descriptors Calculation Speed Experiments: - each experiment was repeated 30 times and the average of the total time needed to complete the calculation for the 10,000 compounds were computed.
* Result and Discussion: - it has several advantages over existing dedicated molecular descriptor calculation software. Firstly, it is free, open source and is licensed as public domain.
* Secondly, PaDEL-Descriptor provides both GUI and command line interfaces, which are available only in CDK Descriptor Calculator GUI, DRAGON and VolSurf.
* Third advantage of PaDEL-Descriptors is that it can work on any platform that support java. It also has Java Web Start technology that is when users make use of this technology to run PaDEL-Descriptor, they will be assured that they are using the most current version.
* The fourth advantages are that it is the only software that supports more than 90 different molecular file formats.
* The last advantage of the PaDEL-Descriptor is its speed, especially in multiple CPU cores environment. Generally, the amount of speedup increases with the number of worker threads.
* The only disadvantage of PaDEL-Descriptor is that it does not calculate as many descriptors as some software like DRAGON, MODEL, Molconn-Z, and PreADMET Descriptor.
* The result showed that 2D descriptors can be calculated at a rate of greater than 20,000 molecules per second but there were also some 2D descriptors that were calculated at a rate of less than 1000 molecules per second.
* Conclusion: - it is a multithreaded software which make full use of the multiple CPU cores in modern desktop computers to increases the speed of calculation of molecular descriptors.