|  |  |  |
| --- | --- | --- |
| **Protists name** | **Disease** | **Symptoms** |
| **Giardia intestinalis (strain ATCC 50803 / WB clone C6)** | Giardiasis | Diarrhea, stomach cramps, nausea, dehydration |
| **Leishmania donovani** | Visceral leishmaniasis (Kala-azar) | Fever, weight loss, enlarged spleen and liver, anemia |
| **Leishmania mexicana** | Cutaneous leishmaniasis | Skin ulcers, lesions, scarring |
| **Plasmodium falciparum** | Malaria (most severe form) | High fever, chills, sweating, organ failure in severe cases |
| **Toxoplasma gondii** | Toxoplasmosis | Flu-like symptoms in healthy individuals, severe complications in immunocompromised people and pregnant women (can affect the fetus) |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | **Organism** |  | | --- | --- | | |  | **Disease** | | --- | --- | | **Symptoms** |
| **Aspergillus flavus** | Aspergillosis, Aflatoxicosis | Lung infections, allergic reactions, liver damage (due to aflatoxins in contaminated food) |
| **Aspergillus fumigatus** | Invasive aspergillosis, Allergic bronchopulmonary aspergillosis (ABPA) | Fever, cough, breathing difficulty, lung infections |
| **Aspergillus parasiticus** | Aflatoxicosis | Liver damage, increased risk of liver cancer (through contaminated food) |
| **Candida albicans** | Candidiasis (oral thrush, vaginal yeast infections, invasive candidiasis) | White patches in mouth/throat, itching, burning, bloodstream infections |
| **Candida glabrata** | Candidiasis | Urinary tract infections, bloodstream infections (in immunocompromised individuals) |
| **Cryptococcus neoformans var. grubii** | Cryptococcosis (primarily affects immunocompromised individuals, such as those with HIV/AIDS) | Pneumonia, meningitis, headache, fever |
| **Paracoccidioides brasiliensis** | Paracoccidioidomycosis | Skin ulcers, lung infections, fever, weight loss |
| **Pneumocystis carinii (now known as Pneumocystis jirovecii in humans)** | Pneumocystis pneumonia (PCP) | Shortness of breath, cough, fever (mostly affects immunocompromised individuals) |

\*

Amino Acid: Serine, Tyrosine, Threonine (abbreviated as S,T,Y)

Link: <https://github.com/Elham-khalili/Soybean-P-sites-Prediction>

Link: <https://github.com/mahdip72/ubi>

Features you need to use:

**1- Amino Acid Composition (AAC)**

**2- Dipeptide Composition (DPC)**

**3- Tripeptide Composition (TPC)**

**4- Position Specific Scoring Matrix (PSSM)**

**5- Physicochemical Properties -Based Feature**

**6- Binary-encoding (BE) :** Previous studies have shown that neighboring AAs of phosphorylated residue are the key sequence-based features for the prediction of phosphorylation sites. Binary-encoding (BE) method was used in which each AA corresponded to a 20- dimensional binary vector comprised of elements, ‘0’ and ‘1’. For example, Alanine is represented as a vector ‘10000000000000000000’, Serine as ‘00000000000000010000’ and so on and thus we obtained a 420-dimensional vector (21 × 20 = 420).

**Algorithm you need to use:**

RF,SVM, LR-L1, LR-L2, XGBoost ,MLP,GBT and TabNet