Cher Monsieur,

Voici quelques questions importantes pouvant m’aider à mieux comprendre l’intention de votre projet et qui mèneront la suite de mon stage et affecteront les méthodes statistiques empruntées. Avec plaisir de les discuter avec vous la prochaine fois.

1. Est-ce que le but de cet examen psychomoteur et la collecte des données durant les années passées s’arrête à hiérarchiser la sensibilité des 26 critères et d’étudier leur importance, ou est-ce qu’en plus de cela, la mise en place d’un barème psychomoteur afin de prédire dans le futur si une personne appartient à un sous-groupe pathologique, uniquement à partir des résultats de ses tests psychomoteurs, est aussi envisagée, voire l’intention la plus importante du projet ? Puisqu’en fonction de ce que l’on veut faire, les méthodes statistiques appliquées peuvent se différer. En effet, dans le cas où la prédiction fait partie de l’objectif de la recherche, je proposerais des méthodes de l’apprentissage supervisé (régression logistique, SVM, forêts aléatoires, kNN, etc.), qui sont assez efficaces dans un problème de classification typique comme le nôtre. On pourrait en discuter ensemble et je pourrais éventuellement faire une petite présentation sur ce genre de méthodes.
2. Toujours dans le cas de la prédiction, est-ce qu’une simplification de critères (diminution du nombre de tests psychomoteurs) est également envisagée ? Si le passage de tous les tests est coûteux ou compliqué, ou encore si l’on veut rendre le résultat de l’examen plus explicable en gardant seulement les tests les plus représentatifs, c’est bien la situation dont on parle. Dans ce cas-là des traitements visant à limiter le nombre de critères peuvent être considérés. En revanche, si le passage et la conservation de tous vos tests ne sont pas compliqués et c’est bien la qualité de la prédiction qui est au premier rang, alors ce n’est pas forcément la peine de faire ça.
3. Comme vous l’aviez certainement aussi remarqué, en général les 26 critères sont de telle façon que souvent pour un certain critère peu de témoins ont obtenu de 0, mais en revanche le nombre de patients avec un résultat 1 est considérable. Cela veut dire que la plupart des critères sont « de mauvaise qualité » en termes de séparation des deux groupes. On aurait probablement du mal à différencier parfaitement les deux catégories en choisissant seulement 3, 4 critères les plus sensibles. De l’autre part, toutefois, si on ne considère que le nombre de témoins qui ont raté un test comme seule mesure de qualité, la plupart des critères sont quand-même très bons. Cela veut dire qu’on ne parle ici plus de la séparation totale, mais d’une « séparation au sens unique » : si une personne a obtenu un 0 pour un critère, on sera « quasiment » sûr qu’il a une grande chance de subir une pathologie en question ; par contre s’il a obtenu un 0, on peut rien y dire. La question qui se pose ici est, si dans la pratique le faux positif (la personne qui se fait diagnostiquer à tort une maladie psychiatre mais ne l’a en fait pas) est plus grave que le faux négatif (la personne qui passe l’examen mais possède en revanche cette maladie), ou si c’est l’inverse, ou c’est également important ? ça aura un impact non seulement sur les méthodes de prédiction, mais déjà sur la hiérarchisation des critères.
4. L’étude principale dans mon stage doit d’après vous se trouver notamment dans le champ Patients vs. Témoins, Schizophrènes vs. Témoins ou Dépressifs vs. Témoins ? Ou encore dans tous les trois si temps nous permet ? Et la détection des sous-catégories de la schizophrénie ça vous intéresserait peut-être aussi (cela exige d’ailleurs plus de données de patients à mon avis) ?
5. En ce qui concerne la cotation binaire : la cotation binaire me gêne en effet pas du tout d’après ma première constatation de vos données. C’est juste que pour faire des analyses statistiques un peu plus avancées sur d’autres sujets intéressants, il vaut mieux avoir plus de données (surtout patients). J’ai cru au début que vous avez tronqué vos données collectionnées durant des années pour vous débarrasser du biais par rapport à l’âge et au sexe. Du coup il ne reste que 71 patients tous en dessous de 40 ans. Mais je viens de me rendre compte que ces 71 patients sont un nouveau jeu de données qui n’ont rien à faire avec l’ancien jeu de données avec lequel vous aviez effectué vos recherches en 2008 et en 2012 ? Et que ce n’est pas facile de transformer les résultats d’une cotation de 0 à 4 en binaire ? Dommage qu’on ne puisse pas profiter des anciennes données.
6. La cotation à 5 niveaux donnera à nos données sans doute une forme plus riche d’un point de vu statistique. De l’autre part je comprends tout à fait la problématique que vous m’aviez expliquée, que cela sera trop subjectif pour que l’étude puisse facilement se généraliser. Si jamais vous y hésitez, j’ai un conseil à vous proposer, c’est qu’on ne transmettrait en fin de compte pas un barème concret ou une formule précise pour qu’un médecin puisse déterminer tout de suite la pathologie de la personne en lisant le résultat des tests de celui-ci, mais plutôt une méthode de calcul que le médecin doit ensuite appliquer à sa propre base de données qu’il a établie au fils de temps, afin que la subjectivité ne joue plus aucun rôle car c’est toujours le même médecin. Cela étant dit, c’est vous qui êtes dans la profession et connaissez l’objectif de votre projet beaucoup mieux que moi, je suis donc à votre écoute à cet effet.
8. D’après ce que j’ai compris dans vos articles des années 2008 et 2012, le but de tout cet examen ne s’arrête pas à hiérarchiser la sensibilité des 26 critères. Par contre il vise plus loin à vraiment mettre en pratique les résultats obtenus, c’est-à-dire de diagnostiquer (ou prédire en terme statistique plus précis), seulement à partir de son bilan des tests psychomoteurs, si un futur patient subit de la schizophrénie ou de la dépression, avec une bonne qualité de prédiction (taux d’erreur bas). Et donc « d’aider l’analyse psychiatre » est aussi dans ce sens-là. Est-ce que vous pouvez me le confirmer ? Puisqu’en fonction de ce que l’on veut faire, les méthodes statistiques peuvent être très différentes. En effet, si le but de votre étude n’est pas seulement de savoir l’importance relative de chaque critère, des méthodes statistiques plus synthétiques et efficaces, qui p. ex. prendront en compte tous les 26 critères en même temps, peuvent s’appliquer. Je reviendrai plus tard sur ce point-là.
9. Vu que la cotation se fait en 0 et 1, la hiérarchisation de la sensibilité des critères est plutôt simple (car il n’y a que 4 possibilités pour chaque critère : nombre de patients qui se font coter 1, nombre de patients qui se font coter 0, nombre de témoins qui se font coter 1, nombre de témoins qui se font coter 0. On n’a pas trop de choses à y faire. Ici on se balade en premier temps dans la question de différencier les patients et les témoins). Cependant, une question pourrait être quand-même intéressante d’être posée, comme dans presque tous les autres diagnostiques médicaux d’ailleurs : est-ce que vous préfériez, concrètement pour cet examen, une erreur de 1er espace à … ou inversement ? C’est-à-dire, dans la future mise en pratique de votre étude, est-ce que la faute qu’une personne est diagnostiquée en tant que patient qui ne l’est en fait pas est plus grave, ou la faute qu’un patient est catégorisé en tant que normal qui subit en fait une pathologie en question ? Cela peut affecter notre jugement sur la « qualité de séparation » (alors la sensibilité) des critères.
10. Dans le fichier Excel que vous m’aviez donné il y a 25 schizophrènes et 19 dépressifs diagnostiqués. C’est-à-dire les autres 27 patients ne sont pas encore cotés. Est-il prévu comme ça ?

Avec ces questions (with this bear in mind) on peut maintenant parler de notre jeu de données.

Comme vous l’aviez certainement aussi remarqué, en général les 26 critères sont de telle façon que souvent les témoins ont obtenu peu de 1, mais en revanche le nombre de patients avec un résultat 0 est considérable. Cela veut dire

1. La plupart des critères sont de mauvaise qualité en termes de séparation des deux groupes. On aurait probablement du mal à différencier « parfaitement » les deux catégories en prenant seulement 3, 4 critères les plus sensibles.
2. Par contre si on ne considère que l’erreur de 1ère espèce (les témoins avec un 1) comme seule mesure de qualité, la plupart des critères sont quand-même très bons. Cela au prix qu’on ne parle ici plus de la séparation totale, mais une « séparation au sens unique » : si une personne a obtenu une 1 pour un critère sensible, on sera « quasiment » sûr qu’il a une grande chance de subir une pathologie en question ; par contre s’il a obtenu un 0, on peut rien y dire. Et ça c’est aussi une des raisons pour lesquelles j’aurais besoin de savoir quel type d’erreur est préféré dans la pratique psychomotrice.

Nous nous trouvons toujours dans le cadre de la recherche d’un bon « classifieur » qui catégorisera une personne selon ses résultats de tests. Cette procédure on l’appelle aussi « sélection de caractéristiques » ou encore « sélection de variables ». Les techniques que vous avez utilisées dans vos articles, d’analyser les critères un par un, font partie des méthodes statistiques classiques des fois aussi appelées « filtre » (parce que vous en effet filtrez les « bons » critères). Celles-là a pourtant ses limites, puisqu’elles ne prennent pas en compte de la relation inter-variable et tendent à choisir des critères redondants. Je vous le montre par un exemple (voir pièce jointe ‘exemple1’). Ici c’est les éléments appartenant à deux différents groupes graphiquement représentés par deux nuages de points (blanc et noir), (en bas à gauche et en haut à droite, les deux graphiques sont le même) avec deux variables (donc deux critères dans notre cas) x (horizontalement) et y (verticalement), les deux histogrammes à côté étant un résumé de tous les points sur resp. x et y. Vous voyez que quand on regarde chacun des critères seul, les points des deux groupes ne sont pas séparables (donc ni x ni y seul n’est sensible). On nous donnera un nouveau point avec ses coordonnées (x, y) sans dire à quel groupe il appartient, nous ne saurons pas trop quel groupe nous devons lui attribuer. Par contre en 2D on trouvera facilement une droite qui passe diagonalement entre les deux groupes et elle peut servir comme un bon classifieur : si le futur point se trouve en dessus de la droite on lui attribuera le groupe noir et vice versa. C’est une méthode synthétique. Vous diriez peut-être que c’est facile et qu’il suffit alors de regarder les graphiques. Mais il arrive des fois très bien que les points ne sont pas séparables en 2D mais en 3D, ou qu’ils ne sont pas séparables en 3D mais en 4D… enfin, nos yeux humains ne peuvent constater que des graphiques jusqu’à 3 dimensions et sont incapables d’accomplir une telle tâche en haute-dimensions.

Heureusement, il y a aussi des méthodes appelées « wrapper » (enveloppe en français) qui sont efficaces dans cette situation. Dans l’exemple précédent, on ne s’amuse alors plus à hiérarchiser les critères selon leur sensibilité. Mais on s’intéresse plutôt à trouver une telle droite à bien séparer les deux nuages de points. Ce genre de méthodes étant devenu assez répandu dans les 10 années passées dans le diagnostic médical comme dans notre cas, il est fort en termes de précision de la prédiction par rapport aux méthodes de filtre, mais au prix d’un certain degré d’explicabilité (on aura à la fin un classifieur genre : calculez-moi 0,2 \* critère1 + 0,3 \* critère2 + 0.4 \* critère3 et je vous dirai si vous êtes malade. Donc on risquera d’avoir mal à expliquer à la personne pourquoi est-elle ou n’est-elle pas malade, mais on peut effectivement être beaucoup plus sûr de la prédiction). A vous de voir si ce genre de méthodes vous intéresse.

Voici une visualisation en 2D puis 3D de nos données par la méthode d’ACP. Elle a été faite de façon que je considère toutes les 26 dimensions des données et je n’en garde que les 2 ou 3 premières directions sur lesquelles les données sont les plus étendues. Ces graphiques illustrent encore l’idée présentée ci-dessus : bien que aucun des critères ne possède une bonne séparabilité, ensemble, ou au moins sur les 3 premières directions, les points des deux groupes se regroupent quand-même entre eux et de façon qu’on puisse y tracer une « frontière ».

Vous trouvez dans un autre dessin les mêmes points mais libellés en « témoins », « schizophrènes », « dépressifs » et « patients non libellés ». Cette fois-ci vous verrez que les dépressifs sont susceptibles de ne pas pouvoir être distingués des schizophrènes, pas avec la cotation en 0 et 1 au moins. Ils sont bien mélangés dans presque la même région. Aussi, je trouve que la tâche de catégoriser les patients en sous-groupes (avec différentes sortes de schizophrénie) pourrait être difficile avec cette cotation et ce nombre d’observation. Toutefois, cela n’étant qu’une première constatation visuelle, il faudra des méthodes rigoureuses pour confirmer ces difficultés.

Cela étant dit, j’aimerais souligner que les méthodes de filtre (ce que vous avez utilisé) sont également d’importance. On ne doit en aucun cas les omettre. Elles servent souvent dans l’étude préliminaire des données à donner une impression générale sur la qualité des variables.

En ce qui concerne la cotation : j’avais bien compris l’explication que vous m’avez donnée la dernière fois pourquoi vous êtes passés d’une cotation à 5 niveaux à une cotation binaire. Etant donné que cette cotation simple peut avoir un impact sur le résultat d’analyse (p. ex. comme vous l’avez vu plus haut, les dépressifs ne se distinguent visuellement pas des schizophrènes, mais si la cotation a un nombre de niveaux plus riche, ce n’est peut-être plus le cas), je pense qu’il y a deux solutions possibles pour contourner la subjectivité :

1. Dans la pratique on stipule des instructions très précises pour que tous les médecins arrivent à peu près à la même cotation pour chaque critère.
2. On ne transmettra pas en fin de compte un barème concret pour qu’un médecin puisse déterminer tout de suite la pathologie de la personne, mais plutôt une méthode de calcul qu’il faudra appliquer à la propre base de données de patients du médecin ou du département, vu que la façon de coter d’un même médecin ne changera pas au fils de temps.

Bien sûr que ce sont